Intervening with Peer Ecology: the impact of Cooperative Learning on Peer Acceptance, Empathy and Pupils' Experiences of Bullying in School

Rachel Faizey

Thesis submitted to the University of Nottingham for the degree of Doctor of Applied Educational Psychologist

May 2019

Table of Contents

Table of Contents 2

List of Tables 9

List of Figures 12

List of Appendices 14

List of Acronyms 15

Abstract 16

Acknowledgements 17

Introduction 18

1 Literature Review 20

1.1 What is CL? 20

1.2 A brief history of CL 21

1.3 Theoretical underpinnings of CL 21

1.3.1 Social interdependence theory 21

1.4 Cognitive-developmental theory 22

1.4.1 Developmental perspective 22

1.4.2 Cognitive elaboration perspective 23

1.4.3 Behavioural-social learning theory 23

1.4.4 Maslow's hierarchy of needs 24

1.4.5 Summary of theoretical perspectives 24

1.5 Models of CL 25

1.5.1 Johnson and Johnson's conceptual approach 'Learning Together' model 25

1.5.2 Slavin's student team learning method 26

1.5.3 Kagan's structural approach 26

1.5.4 A critical review of CL models 27

1.5.5 Social skills 28
1.5.6 Group composition 29
1.5.7 Implementation 30
1.6 Reported benefits of CL 31
1.6.1 Academic outcomes 32
1.6.2 Social outcomes 33
1.7 Classroom peer ecology 34
1.8 PA, empathy and bullying 34
1.8.1 PA 34
1.8.2 Theoretical frameworks and PA 35
1.8.3 PA and CL 36
1.8.4 Empathy 37
1.8.5 Empathy and CL 38
1.9 Bullying 39
1.9.1 Bullying and CL 40
1.10 A joining of cooperative learning, empathy, PA and bullying 42
1.11 Systematic literature review 43
1.11.1 Review Question 43
1.11.2 Criteria for considering studies for the review (Inclusion/Exclusion criteria). 44
1.11.3 The search strategy 44
1.11.4 Studies included in the systematic literature review 45
1.11.5 Appraisal of study quality and relevance 54
1.11.6 Synthesis of findings 59
1.11.7 Sample 59
1.11.8 Research design 60
1.11.9 Intervention 60

1.11.10 Measures and dependent variables 63
1.11.11 Strengths and limitations of the systematic review 68
1.11.12 Summary and conclusions of the systematic review 69
1.11.13 Unpublished literature 70
1.12 The rationale for current research 71
1.13 Research questions and hypotheses 72
2 Methodology 73
2.1 Approach to research 73
2.1.1 Positivism 74
2.1.2 Post-positivism 74
2.1.3 Interpretive/constructivist 74
2.1.4 Pragmatic 75
2.2 The paradigm of the proposed research 75
2.3 Research design 75
2.3.1 Research design selected 75
2.3.2 Quasi-experimental designs 76
2.3.3 Limitations of quasi-experimental designs 76
2.3.4 Alternative designs considered 76
2.4 Stakeholder involvement 77
2.4.1 Selection of schools 77
2.4.2 School characteristics 78
2.5 Participants 79
2.5.1 Sampling strategy 79
2.5.2 Class teacher characteristics 79
2.5.3 Class characteristics 80
2.6 Procedure 82
2.6.1 Pilot 82

2.6.2 Training 82
2.6.3 Coaching sessions 83
2.6.4 Structure of the coaching sessions 84
2.6.5 Cooperative learning intervention 84
2.6.6 Comparison 85
2.6.7 Intervention fidelity 86
2.6.8 Diffusion of treatment 87
2.7 Measures 88
2.7.1 Measuring peer acceptance: The Social Inclusion Survey (SIS) 88
2.7.2 Empathy measure 90
2.7.3 Measuring the experience of bullying in school 91
2.8 Data collection procedures 93
2.8.1 Procedure for administering measures 93
2.8.2 Anecdotal Data 94
2.9 Ethical considerations 94
2.10 Reliability and validity 95
2.10.1 Internal validity 95
2.10.2 External validity 96
2.10.3 Issues of reliability 96
3 Results 98
3.1 Parametric and non-parametric tests 98
3.1.1 Outliers 98
3.2 Test Selection 99
3.2.1 Parametric test selection 99
3.2.2 Non-parametric Selection 99
3.3 Effect size 100
3.4 Power 101

3.5 PA 102
3.5.1 Grouping data from multiple schools 102
3.6 Play Same Gender (PSG) 103
3.6.1 Main analysis 103
3.6.2 Main statistical analysis 104
3.6.3 Analysis by individual class groups 105
3.7 Play different gender (PDG) 108
3.7.1 Main analysis 108
3.7.2 Analysis by individual classes groups 110
3.8 Work same gender (WSG) 113
3.8.1 Main analysis 113
3.8.2 Analysis by individual class groups 115
3.9 Work different gender (WDG) 118
3.9.1 Main analysis 118
3.10 Analysis with individual classes 120
3.10.1 Descriptive statistics 120
3.10.2 Statistical analysis for individual class groups 121
3.10.3 PA: Summary of findings 123
3.11 PA: A differential impact 124
3.11.1 Peer same gender 125
3.11.2 Statistical analysis 126
3.12 Play different gender 128
3.12.1 Statistical analysis 129
3.13 Work same gender (WSG) 131
3.13.1 Statistical analysis 132
3.14 Work different gender 133
3.14.1 Statistical analysis 134

3.15 Empathy 135
3.15.1 Descriptive Statistics 135
3.15.2 Empathy: Summary of Findings 138
3.16 Bullying 139
3.16.1 Bullying index 139
3.16.2 Descriptive Statistics 139
3.16.3 Statistical Analysis 140
3.16.4 General Aggression Index 143
3.16.5 Analysis by individual class 145
3.17 Summary of Findings 147
3.18 Intervention Fidelity 151
3.18.1 Adherence 151
3.18.2 Dosage 151
3.18.3 Intervention Fidelity in individual schools 151
3.18.4 Summary of Intervention Fidelity 155
3.18.5 Anecdotal data 156
3.19 Conclusion 160
4 Discussion 162
4.1 Research findings 162
4.1.1 Research Question One: What is the impact of CL on pupils' PA? .. 162
4.1.2 Research Question Two: Does the impact of CL on PA differ depending on pupils' initial levels of PA? 163
4.1.3 Links to the literature: PA 163
4.1.4 Research Question Three: what the impact of CL on pupils' empathy is 165
4.1.5 Links to the literature: Empathy 165
4.1.6 Research question four: What is the impact of CL on pupil's experience of bullying in school. 166

4.1.7 Links to the Literature: Pupils experiences of bullying in school 167
4.1.8 Procedural Aspects 168
4.2 Methodological Review 168
4.2.1 Internal Validity 168
4.2.2 Regression to the mean 169
4.2.3 Intervention fidelity 169
4.2.4 External validity 170
4.2.5 Review of measures 171
4.2.6 Review of research design aspects 172
4.3 Implications of the findings 173
4.3.1 Implications for practice 173
4.3.2 Implications for future research 174
4.4 Unique contribution of the study 178
4.5 Summary and Conclusions 178
5 References 180
6 Appendices 201

List of Tables

Table 1. 1 Summary of key information extracted from studies..... 47
Table 1. 2 Outline of the Weight of Evidence Criteria used within the research.. 54
Table 1. 3 Weight of Evidence analysis for the studies included in the systemic review 55
Table 1. 4 Author, Title and weight of evidence for studies included in the systematic review..... 58
Table 1. 5 Summary of variables and measures used in each study included in the review 65

Table 2. 1 Key characteristics of the two participating schools..... 78
Table 2. 2 Class teacher characteristics (intervention and control)..... 80
Table 2. 3 A table to show the class characteristics for each school and condition providing details as to the total number of pupils in each class, gender and number of pupils with and without consent..... 80
Table 2. 4 The total number of participants in intervention and control conditions, including information on consent levels. 81
Table 2. 5 Intervention adherence: raw scores and percentages in waitlist control condition 87

Table 3. 1 Effect Sizes and Descriptors for Cohens d and Pearson's r. 100
Table 3. 2 A table to show the Mean, Standard deviation and Median for pre and post-test scores and gain scores for PA measure in the Play Same Gender (PSG) Context. 103
Table 3. 3 A table to show the Mean, Standard deviation and median for pre and post-test scores and gain scores for the PA measure Play Same Gender each class. 105
Table 3. 4 A Table to show results from statistical analysis for each class for PA measure for Play Same Gender (PSG). 107
Table 3. 5 A table to show the Mean, Standard deviation and Median for pre and post-test scores and gain scores for the Play Same Gender (PDG) Context. 108

Table 3. 6 A table to show the Mean, Standard deviation and median for pre and post-test scores and gain scores for Play Same Gender Context in each clas..... 110
Table 3. 7 A Table to show results from statistical analysis for each class for PA measure for Play Different Gender (PDG)..... 112
Table 3. 8 A table to show the Mean, Standard deviation and Median for pre and post-test scores and gain scores for the Work Same Gender (WSG) Context..... 113
Table 3. 9 A table to show the Mean, Standard deviation and median for pre and post-test scores and gain scores for Work Same Gender (WSG) context in each class. 115
Table 3. 10 A Table to show results from statistical analysis for each school and class for Work Same Gender (WSG)..... 117
Table 3. 11 A table to show the Mean, Standard deviation for pre and post-test scores and gain scores for the Work Different Gender (WDG) context..... 118
Table 3. 12 A table to show the Mean, Standard deviation and median for pre and post-test scores and gain scores for Play Same Gender in each class..... 120
Table 3. 13 A Table to show results from statistical analysis for each school and class for Work Different Gender (WDG). 122
Table 3. 14 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in Play Same Gender (PSG) context..... 125
Table 3. 15 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in Play Same Gender (PDG) context. 128
Table 3. 16 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in Work Same Gender (WSG) context. 131
Table 3. 17 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in Work Same Gender (WSG) context. 133
Table 3. 18 A table to show Mean, Standard deviation and gain scores pre-test and post-test on total empathy scores. 135
Table 3. 19 A table to show Mean, Standard deviation and gain scores pre-test and post-test on total empathy scores for each class..... 137

Table 3. 20 A table to show the mean and median for Bullying Index gain scores pre and post-test for the intervention and control group. 139
Table 3. 21 A table to show the Bullying Index and mean gain scores for pre and post-test for the intervention and control group in each class. 141
Table 3. 22 A table to show mean and median for the General Aggression Index and gain scores pre and post-test for intervention and control group. 143
Table 3. 23 A table to show General Aggression Index and mean gain scores at pre-test and post-test for intervention and control group for each class. 145
Table 3. 24 A table to show a summary of findings, both overall and individual classes for each research question 147
Table 3. 25 A table to show raw and average adherence score in percentages across four fidelity checks for each class. 153
Table 3. 26 A table to show a summary of the information relation to intervention dosage including duration, total number of CL sessions, duration of CL structures, time spent using in CL (in minutes) and the structures used. 154

Table 4. 1 Possible Suggestions for future research 176

List of Figures

Figure 1. 1 A flow diagram summarises the search process 46

Figure 3. 1 A bar chart to show scores at pre and post-test for the intervention and control group for PSG. 104
Figure 3. 2 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for PSG.. 106
Figure 3. 3 A bar chart to show scores at pre and post-test for the intervention and control group for PDG. 109
Figure 3. 4 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for PDG. 111
Figure 3. 5 A bar chart to show scores at pre and post-test with confidence intervals for the intervention and control group for WSG. 114
Figure 3. 6 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for WSG. 116
Figure 3. 7 A bar chart to show scores at pre and post-test with confidence intervals for the intervention and control group for PSG. 119
Figure 3. 8 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for WDG. 121
Figure 3. 9 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for Play Same Gender (PSG) context. 126
Figure 3. 10 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for PDG. 129
Figure 3. 11 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for Work Same Gender (WSG) context. 132
Figure 3. 12 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for Word Different Gender (WDG) context. 134
Figure 3. 13 A bar chart to show mean pre and post-test total empathy scores for intervention and control group. 136
Figure 3. 14 A bar chart to show mean pre and post-test total empathy scores for intervention and control group in each class. 138
Figure 3. 15 A bar chart to show the Bullying Index and mean gain scores at pre and post-test for the intervention and control group. 140

Figure 3. 16 A graph to show the Bullying Index and mean gain scores at pre and post-test for intervention and control groups in each class. 142
Figure 3. 17 A graph to show the General Aggression Index and mean gain scores at pre and post-test for intervention and control groups. 144
Figure 3. 18 A graph to show the General Aggression Index and mean gain scores at pre and post-test for intervention and control groups in each class. 146

List of Appendices

Appendix A A table depicting the source and number of results obtained from the systematic search 201
Appendix B Summary of Full text studies excluded from synthesis 203
Appendix C A table summarising the Weight of Evidence Criteria 205
Appendix D A table illustrating a list of studies included and excluded from the unpublished articles systematic review search 207
Appendix E Information provided to school for research recruitment 209
Appendix F Information Sheet for Class Teachers 213
Appendix G Parental Information Sheet and consent forms 216
Appendix H Example of Teacher Resource Booklet 219
Appendix I Coaching Session Structure 224
Appendix J Intervention Fidelity Checklist 226
Appendix K Social Inclusion Survey 229
Appendix L My Life in Schools Checklist 231
Appendix M Empathy Measure 232
Appendix N Ethics amendment letter 233
Appendix O Pupil Information 235
Appendix P Pupil Consent Form 240
Appendix Q Ethical approval from the University of Nottingham Ethics Committee 241
Appendix R Parent debriefing letter 243
Appendix S Examples of completed teacher record booklets 245
Appendix T Bullying Index Data 247
Appendix U General Aggression Index Data 249

List of Acronyms

Acronym	Description
CL	Cooperative Learning
EP	Educational Psychologist
EPS	Educational Psychology Service
LA	Local Authority
PA	Peer Acceptance
PSG	Play Same Gender
PDG	Play Different Gender
STAD	Student Teams-Achievement Divisions
SIS	Social Inclusion Survey
TEP	Trainee Educational Psychologist
WDG	Work different Gender
WoE	Weight of Evidence
WSG	Work Same Gender

Abstract

Cooperative Learning (CL) is a peer-mediated, instructional intervention with students working together to accomplish shared learning goals (Johnson & Johnson, 2000). The study aimed to build upon existing CL and Peer Acceptance (PA) research and extend this by looking more specifically at whether CL, through the promotion of positive interdependence, could have beneficial outcomes for bullying and associated factors such as empathy.

A quasi-experimental pre-test post-test non-equivalent groups design was used. Data was collected from three classes in two schools (intervention n=78; control n=79). Teachers delivered CL in Year 4/5 classes over an 11-week period, which included one week of social skills training. The Social Inclusion Survey (SIS) was used to measure PA (Fredrickson and Furnham, 1998). Empathy was measured using an adapted questionnaire from the Emotional Literacy Scales (Faupel, 2003) and the My Life in School Checklist was used to measure pupils' experience of bullying (Arora & Thompson, 1987).

No overall significant differences were found between the intervention and control group for PA in work and play contexts across same and different genders. Although, there were significant differences between the intervention and control group in some individual classes, this represented a decrease in overall PA. There also appeared to be a differential impact, in some contexts, in favour of pupils with initially low levels of PA.

The study highlighted difficulties associated with measuring empathy and provided some justification for further research to explore the links between CL and bullying, with anecdotal staff data suggesting a decrease in bullying and formal data indicating that CL may have at least a stabilising effect on bullying.

Findings varied across individual classes and schools, indicating caution needs to be taken when interpreting results. The study highlights the importance of intervention fidelity, illustrating the complexity involved in understanding implementation aspects, in evidence-informed approaches.

Acknowledgements

I would like to express by utmost gratitude to the following people for their support during the completion of this research and doctoral training.

- The schools, staff and pupils who took part in the research for their participation and contributions.

- The tutor team at the University of Nottingham, particularly my supervisor Anthea Gulliford, for her guidance and unwavering support.

- My fellow Trainees for their friendship and support throughout the doctorate.

- Family and friends for their patience and endless encouragement.

Introduction

During professional training the author became aware of Cooperative Learning (CL) as a pedagogical approach and its associated evidence highlighting both social and academic benefits. CL has been defined as a peer-mediated, instructional intervention in which students work together to accomplish shared learning goals (Johnson & Johnson, 2000), with the principles of positive interdependence and individual accountability distinguishing CL from generic group work.

The potential significance of CL comes to fruition when thinking about the societal roles of schools. Schools could be viewed as a microcosm of society where children learn about rules, relationships and interactions which will stand them in good stead for their future lives. What struck the author was that despite evidence for CL it is not often used within schools. The UK education system seemingly values individualist or competitive learning (Baines, Blatchford, & Kutnick, 2003) with success in the classroom often contingent on beating others, with certain pupils having to fail for others to be successful. Although competition has a place throughout life it can be questioned whether we are teaching children the skills to function cooperatively later in life, which can be fundamental to adulthood as illustrated in the following quote:

> *'We are exposed to competition in every aspect of our lives; it pervades the home, the classroom and the media alike, and is frequently a source of stress, particularly with children who quickly see themselves as failures. Yet it is cooperation that has always been an intrinsic part of human nature'.*
>
> (Mashender, 1986, p.5)

If cooperation is an intrinsic part of human nature, this begs the question as to why we are not teaching children how to cooperate as part of the curriculum. Being able to work and cooperate with others is not only seen as pivotal to human functioning, but group work has been found to be one of the most important elements in interventions associated with a decrease in bullying and social isolation (Ttofi & Farrington, 2011). Bullying is a phenomenon which remains of great significance and the importance of tackling bullying is reiterated in the recent green paper on transforming young people's mental provision due to its impact on the mental health of pupils (Department for Health, 2017). Consequently, the government is providing £1.6 million worth of

funding into initiatives to tackle bullying. Despite the focus on decreasing bullying in schools, research into anti-bullying interventions has often demonstrated small effect sizes overall (Tfofi & Farrington, 2011).

How students perceive each other and interact with one another is typically a neglected aspect of instruction (Johnson & Johnson, 1992). Suggestions in the literature advocating CL as a potential approach to address bullying (Hymel, McClure, Miller, Shumka, & Trach, 2015) and costly interventions showing limited effect sizes prompted the author to explore whether CL could play a role in reducing bullying. The author felt the impact of CL on pupils' experiences of bullying and associated factors such as Peer Acceptance (PA) and empathy was worthy of further investigation. If CL could potentially facilitate improvements in factors associated with bullying and experiences of bullying itself the question arises as to whether CL could serve as a potential preventative intervention, without the need for extensive training, materials, significant time commitment or withdrawal groups.

Although the evidence base regarding CL shows potential in terms of positive outcomes, to the author's knowledge explicit investigation of the role CL could play in bullying is limited. Additionally, the evidence base indicating potential benefits lay primarily outside of the UK with the focus often being on special populations. Two doctoral theses, Craig (2010) and Smith (2017), have lay foundations for research in the UK highlighting some positive impacts of CL on PA and peer support. The aim of this research was to build upon existing research and extend this by making a unique contribution, looking more specifically at whether CL could have beneficial outcomes into bullying and associated factors such as empathy.

1 Literature Review

In order to provide the conceptual foundations for this study, this chapter outlines the theoretical and empirical basis for CL with reference to relevant literature along with links to PA, empathy and bullying. This is followed by a systematic review which aims to gain a better understanding of the current evidence base linking CL, PA, empathy and bullying.

1.1 What is CL?

CL describes a peer-mediated, instructional intervention in which students work together to accomplish shared learning goals (Johnson & Johnson, 2000). Instructional environments are typically organised in three ways; cooperative, competitive and individualistic (Johnson & Johnson, 1991). In CL environments, students work collectively towards a group goal, whereas in competitive environments students work against one another to see who is best at a task (Marr, 1997). In individualistic environments, the emphasis is on independent work with goals being independent of other students (Marr, 1997).

All three types of instruction exist within a classroom. However CL, compared to competitive and individualistic environments, is thought to promote academic and prosocial development for all learners (Johnson, & Johnson, 1991). Early research supported this view as children with disabilities were more likely to be chosen as friends in hetreogenous CL groups compared to individualistic learning (Armstrong, Johnson, & Balow, 1981). With Regard to academic development Skon, Johnson, & Johnson (1981) found cooperative interaction, compared with competitive and Individualistic learning, promoted higher achievement and greater cognitive reasoning strategies.

Recent research exploring the benefits of comparing CL to different types of instruction appears limited, with Johnson's work dominating the literature. Although recent research has shown significant differences in achievement scores, favouring CL conditions (Alabekee, Egbulefu, Samuel, & Osaat, 2015), along with greater enjoyment and a stronger mastery goal orientation (Plass et al., 2013).

1.2 A brief history of CL

CL is not a new idea (Marr, 1997). In the late 1700s, Joseph Lancaster took the idea of CL groups to America, opening the Lancastrian school in 1806. In the 19th century, Colonel Francis Parker advocated for CL, and instructional methods promoting cooperation dominated American education (Johnson et al., 1991). In the 1940s, Deutsch proposed a theory of cooperative and competitive learning which seemed to serve as the foundation upon which further research was built (Deutsch, 1949). However, several theoretical perspectives have been proposed and there remains controversy about why and how CL promotes achievement and social outcomes (Slavin, 2014). Five different theoretical perspectives associated with CL will now be discussed.

1.3 Theoretical underpinnings of CL

1.3.1 Social interdependence theory

Social interdependence theory was first introduced by Deutsch (1949) building upon the work of Lewin (1935). Social interdependence is described as existing when the outcomes of individuals are affected by the actions of others (Johnson & Johnson, 2005). Deutsch (1949) conceptualised two types of social interdependence; positive and negative.

Positive interdependence exists when the actions of individuals promote the achievement of joint goals, in contrast to negative interdependence when the actions of individuals obstruct the achievements of others (Johnson & Johnson, 2009). Positive interdependence, which is proposed as one of the critical elements of CL, is thought to result in promotive interaction. Promotive interaction occurs when individuals encourage and help each other to accomplish the group's goals and achieve success.

Promotive interaction is characterised by students providing each other with the help they need, sharing resources, providing feedback on task and teamwork, challenging conclusions and working constructively to gain mutual goals (Gillies, 2016). This is thought to result in enhanced cognitive processes such as being able to give oral explanations of problems and connecting present with past learning (Johnson, Johnson & Smith, 2013). Promotive interaction is also thought to promote social processes such

as modelling appropriate use of social skills, supporting and encouraging efforts to learn and celebrating success (Johnson, Johnson & Smith, 2013). This is thought to lead to group members feeling more accepted and valued and more willing to help others in return (Gillies, 2016).

On the other hand, negative interdependence and competition occur when only one or a few individuals can reach a goal, promoting oppositional interaction. When oppositional interactions occur individuals discourage, block and obstruct each other's efforts to achieve their goals (Johnson & Johnson, 2009). Deutsch (1949) also described situations where there is an absence of goal interdependence, whereby individuals work independently with achievements unrelated to the effort and achievement of peers (Johnson & Johnson, 2009).

CL is thought to be effective because it promotes positive interdependence and promotive interaction (Johnson & Johnson, 2009), which are thought to motivate students to re-evaluate the social desirability of others (Bierman, 2004). On the other hand, it is suggested that positive interdependence might not be enough. Hymel et al. (2015) suggest that it may be possible for students to work towards a shared goal without necessarily developing supportive relationships. Cognitive Developmental Theory provides an alternative theory for the success of CL.

1.4 Cognitive-developmental theory

1.4.1 Developmental perspective

The developmental perspective assumes that interaction among children around appropriate tasks increases their mastery of critical concepts (Slavin, 2015). This perspective draws upon the work of Vygotsky (1978) and his concept of the Zone of Proximal Development which is:

'The distance between the actual development level determined by independent problem solving, and the level of potential development that can be achieved through adult guidance or in collaboration with more capable peers' (Vygotsky, 1978, p86).

Vygotsky (1978) asserted that certain developmental processes are only able to operate when the child is interacting with people in their environment and in cooperation with others. In this view, collaboration with peers promotes growth because children of

similar ages are likely to be operating in another's Zone of Proximal Development. When children work cooperatively peers often provide information, prompts, reminders and encouragement to requests for help or their perceived need for help (Gillies, 2003). Children are also thought to be well placed to give explanations as they share similar language and vocabulary and can explain concepts that can be easily understood (Webb & Farivar, 1994). Similarly, Piaget (1926) proposed that language, values, rules and social-arbitrary knowledge can only be learnt in interactions with others.

From the developmental perspective opportunities for students to discuss, argue, present and hear one another's viewpoints are critical elements of CL in terms of student achievement (Slavin, 2015). However, despite theoretical support, there is little evidence that CL methods which depend exclusively on interaction produce achievement gains as it is likely cognitive processes act as mediating variables (Slavin, 1995). These processes are acknowledged in the cognitive elaboration perspective.

1.4.2 Cognitive elaboration perspective

Research in cognitive psychology proposes that if the information is to be retained, the learner must engage in some form of cognitive restructuring or elaboration of the material (Callender & Mcdaniel, 2009). One effective means of elaboration proposed is the explanation of the material to someone else, with research on peer tutoring providing evidence for this (e.g. Calhoon, Otaiba, Cihak, King, & Avalos, 2007). There is some support for this perspective, with research highlighting students in CL environments use elaboration and metacognition strategies more frequently (Larson, Dansereau, O'Donnell, Lambiotte, & Rocklin, 1985). Additionally, research on peer tutoring has shown that students who gained the most from CL provided elaborated explanations to others (Webb, 2009). Conversely, it could be argued that the cognitive elaboration perspective does not, on its own, provide an adequate explanation of some of the social outcomes found as a result of CL. Behavioural-social learning theory does, however, attempt to provide some explanation for the positive social outcomes associated with CL.

1.4.3 Behavioural-social learning theory

A classic behavioural learning approach based on the work of Skinner (1953) proposes that when a behaviour is followed by a reward, it is more likely to be repeated. It is

argued CL dramatically increases the immediacy, frequency and desirability of rewards for achievement (Kagan & Kagan, 2015). Motivational perspectives also emphasise that group rewards are essential to the effectiveness of CL (Johnson & Johnson, 2009). CL is considered to be effective because it encompasses modelling which, in accordance with Bandura's (1971) social learning theory, is thought to be one of the most powerful forces for learning as pupils look to emulate successful individuals (Kagan & Kagan, 2015). Nevertheless, it could be simplistic to assume changes will take place purely through the mechanism of rewards and modelling. In contrast, humanistic theories would argue that before learning can occur there is a lower hierarchy of needs which need to be met.

1.4.4 Maslow's hierarchy of needs

Maslow's 'hierachy of needs' (1970) is based on the idea that people attempt to fill deficiencies lower in the hierarchy before we are free to focus on the higher needs. If students do not feel safe and included, their energy will be directed to meeting those needs. Through CL the need for safety is satisfied through social norms and the need for inclusion is satisfied through teambuilding and class building. With these needs met students are free to move up the hierarchy and strive for esteem and knowledge (Kagan & Kagan, 2015).

1.4.5 Summary of theoretical perspectives

Although each theoretical perspective is thought to emphasise the role of reciprocal interaction among peers in constructing knowledge (Tran, 2013), there are fundamental differences in the basic assumptions held within each perspective. Social interdependence theory highlights the importance of group members having joint aspirations and goals, whereas behavioural-social learning theories suggest extrinsic motivators and social reinforcers are what contribute to cooperation. Contrastingly, cognitive-developmental perspectives focus on cognitive and social factors which promote learning and a needs perspective emphasises basic needs being met allowing for further learning and development.

Slavin (2015) attempted to offer a cohesive model of the relationship among the different variables involved in CL, suggesting that group goals might affect cognitive processes directly by motivating students to engage in peer modelling or cognitive elaboration with each other. These group goals may also lead to increased

cohesiveness and an enhanced sense of caring amongst group members, increasing the sense of responsibility towards others achievement (Slavin, 2004). Slavin (2004) further argued that perspectives should be viewed as complementary as each perspective contributes to a more complex understanding of how CL affects student achievement and interpersonal skills.

In addition to different theoretical frameworks CL has also been conceptualised in different ways, with three main models offering different ways to operationalise and implement CL.

1.5 Models of CL

1.5.1 Johnson and Johnson's conceptual approach 'Learning Together' model

Johnson and Johnson's conceptual approach is based on a theoretical framework providing general principles on how to structure cooperative activities. It is not tied to a specific subject area or curriculum, meaning it can be adapted for differences in age and ability (Johnson, Johnson & Johnson Holubec, 1991). Johnson and Johnson propose 5 basic elements fundamental to CL:

- Positive interdependence
- Face to face promotive interaction
- Individual accountability
- Interpersonal and group skills
- Group processing

Positive interdependence occurs when students perceive they need each other to complete the group task and groups need to ensure all members learn the assigned material (Johnson & Johnson, 1992). This can be structured by establishing mutual goals, joint rewards, shared resources and assigned roles (Johnson, Johnson & Holubec, 1991). Individual accountability ensures each member of the group fairly contributes, helping to avoid redundant efforts by group members ensuring each group member is responsible for the final outcome (Johnson & Johnson, 1992). This can be structured through group and individual performance (Johnson, Johnson & Holubec, 1991). Face to Face promotive interaction involves individuals encouraging and facilitating each other to achieve and complete tasks and reach group goals (Johnson

& Johnson, 1992). Small group skills involve children being taught the necessary skills to work effectively in a group situation (Murphy et al., 2005). Johnson & Johnson (1992) emphasise this is an essential element of CL and in order to achieve group goals members must communicate unambiguously, accept and support each other and resolve conflict constructively. The final component, group processing, occurs when groups have a specific time to discuss how well they are achieving their goals and maintaining effective working relationships among members (Johnson, Johnson & Holubec, 1991).

1.5.2 Slavin's student team learning method

Similar to Johnson and Johnson's approach, Slavin's model includes individual accountability as an essential component. However, Slavin's model also emphasises the need for team rewards, which is not an element shared by other CL models, and equal opportunities for success. Slavin's approach involves creating a number of teams who then compete against each other using team goals and team success (Murphy et al., 2005). Slavin emphasises different approaches such as Teams-Games Tournaments, Jigsaw and Student Teams-Achievement Divisions (STAD) (Slavin, 1991). STAD is the simplest form of student team learning methods, which involves students assigned to four or five-member learning teams. Each team is made up of high, average and low performing students. The teacher then introduces new material and the team members must work together to master it. Following team practice, students take quizzes on the material but teammates are not allowed to help each other, the quizzes are then scored and the individual scores are formed into team scores (Slavin, 1991).

1.5.3 Kagan's structural approach

The structural approach to CL shares some concepts with the above models but is based on the application of structures which are content-free ways of organising social interactions in the classroom (Kagan, 1989). Kagan's model contains four basic principles, symbolised using the acronym PIES, which stands for

- Positive interdependence
- Individual accountability
- Equal participation

- Simultaneous interaction

The principles of positive interdependence and individual accountability are aspects which can be found in other models. Kagan & Kagan (2015) propose that typical classrooms do not promote positive interdependence. For example, if a teacher asks a question, many students raise their hand, however the gain of the student selected is a loss for another and its only if the answering student fails that the other students win an opportunity to be called upon. Kagan & Kagan (2015) suggest that this situation inadvertently sets the students against each other and undermines desired outcomes such as sharing, helping, and empathy.

Kagan and Kagan (2015) emphasise the importance of individual accountability maintaining that by teachers calling on volunteers any student can avoid being called upon by simply not raising their hand. Students know there is no individual accountability, therefore many do not put in their best effort or be engaged.

Kagan refers to equal participation which emphasises that everyone makes an equal contribution to the group. Kagan & Kagan (2015) reasoned that only a small number of children will always raise their hand and as you move up the year groups this principle is often violated. Additionally, in a whole class question and answer session, only one student at a time is called upon with most of the class not actively involved. This goes against the principle of simultaneous interaction, which emphasises that as many children as possible should be actively involved at once. The importance of social skills training is also highlighted in the Kagan approach. The approach makes it clear that it is the PIES principles that distinguish CL from group work. If any of the PIES principles are not present then it is not CL and gains cannot be guaranteed (Kagan & Kagan, 2015).

1.5.4 A critical review of CL models

The CL methods outlined differ but are not mutually exclusive (Sharan, 2010). Within the literature, there is a general consensus that group goals and individual accountability are essential components of CL (Slavin, 1990). However, methods of CL are often described as being on a continuum from direct to conceptual (Johnson, Johnson, & Stanne, 2000), varying on the extent to which they structure teaching and freedom for how and what pupils learn. Conceptual CL models such as Johnson & Johnson (1992) provide a framework which can be used as a template when structuring

lessons. Johnson and Johnson's method emphasises social skills and interpersonal communication (Sharan, 2010). In comparison, Slavin's model emphasises mastery of knowledge and motivation (Sharan, 2010). Although the conceptual approach offers more flexibility in its application, autonomy and creativity it is questioned as to whether this flexibility, without a solid understanding of the basic principles, may lead to modifications not reflective of the authentic CL procedure (Sharan, 2010).

Direct models such as Kagan & Kagan (2015), consisting of specific and well-defined techniques, are considered to be extremely helpful in the implementation of CL (Sharan, 2010). The Kagan approach is proposed to be advantageous because it relies on simple structures, takes no special materials or preparation and no change in lesson design or content (Kagan & Kagan, 2015). According to Kagan & Kagan (2015) their approach is the most widely trained and most widely used approach to CL worldwide.

There are subtle differences between each of these models, such as the role of rewards and the amount and level of structure provided. A similarity can be found between Johnson and Johnson's model and Kagan's model as both advocate the use of small group skills when implementing CL, the use of which will now be discussed.

1.5.5 Social skills

CL can take on a structured or unstructured approach with a structured approach incorporating the explicit teaching of social interaction skills (Murphy, Grey, & Honan, 2005). Goodwin (1999) argues that although CL can be successful with students of all ages, learning styles and backgrounds students who have never been taught the prerequisite social skills cannot be expected to work together effectively. Johnson et al. (2000) argue that placing socially unskilled students in a learning group and telling them to work cooperatively will not be successful as students must be taught the small group skills needed for high-quality cooperation (Johnson & Johnson, 1989). Children cooperate and work better when they have the interpersonal and small group skills they need to be able to manage group interactions (Gillies, 2016). Skills such as active listening, considering other ideas, constructively critiquing the ideas of others, sharing resources and taking turns have been suggested as key small group skills (Gillies, 2016). Johnson and Johnson (1989) proposed teachers can follow a series of steps in teaching students interpersonal and small group skills which includes:

1. Students seeing the need to use the skill.

2. Students understanding what the skill is and when it should be used, which is most effectively conveyed through modelling.

3. Students practising the skill again and again.

4. Students processing how frequently and how well they are using the skill.

5. Students persevering in practising the skill.

The effects of a structured versus unstructured approach have been explored empirically in Australian primary schools (Gillies & Ashman, 2000). Findings suggested students in the structured condition performed significantly better in an academic test than the unstructured group, although social outcomes were not included in this study. Additionally, whilst groups need to be structured to embed the key elements of CL and social skills, consideration also needs to be given to the composition of the group.

1.5.6 Group composition

Within the CL literature, there is a consensus that pupils are more likely to achieve when they work in small heterogeneous groups with different abilities, genders and ethnicities. In terms of group size, research which examined the effects of within class groups has shown pupils work better when they worked in groups of 3-4 rather than larger groups (Lou et al., 1996). The use of heterogeneous groups is promoted because it is thought to maximise opportunities for peer support and improves relations across gender and ethnicities ensuring that at least one student in the group can complete the work (Kagan & Kagan, 2015).

The use of heterogeneous groupings has been found to have positive benefits for some students. For example, Lou et al (1996) found low achieving students learned significantly more in heterogeneous groups than homogenous ability groups. Despite these positive outcomes, critics have argued this reasoning may be detrimental to high-ability students (see Robinson, 1990). The use of heterogeneous groups in CL approaches has also faced some objection from the parents of higher achieving students (e.g. Davison, Galbraith, & McQueen, 2008). Evidence, however, indicates that CL is not detrimental to high achieving students, with high achieving students

performing comparably with no significant differences in achievement (Lou et al., 1996).

In fact, evidence seems to show that high ability students may benefit more than low ability students (Terwel, Gilles, van den Eeden, & Hoek, 1999) with higher ability peers benefitting from having to reorganise their own knowledge to be able to explain it to their less-able peers (Gillies, 2003). This evidence is in line with the cognitive elaboration perspective (Callender & Mcdaniel, 2009). Despite these potential benefits for a diverse range of pupils, translating the potential of CL into practice can be more complicated (Sharan, 2010) and is likely to be contingent on its implementation.

1.5.7 Implementation

Even with solid foundations in theory and empirical research, interventions can still yield poor results due to poor implementation (Moir, 2018). This focus on implementation is reflected within implementation science which is the study of how evidence-based programmes can be embedded to maximise successful outcomes (Kelly & Perkins, 2012). Implementation Science is underpinned by theories of organisational change and systems theory (Moir, 2018).

The teacher is critical to the successful implementation of CL. However, it is suggested that once formal training has ceased CL is often abandoned or diluted (Sharan, 2010). Sharan (2010) suggests one of the reasons for these difficulties in embedding CL may be due to the variety of methods and models available, which may be a source of confusion for teachers. As well as potential confusion as to which model of CL to adopt there may be challenges in terms of moving away from more traditional methods of instruction. Buchs, Filippou, Pulfrey, & Volpé (2017) highlighted that the most frequent instructional strategies reported are traditional ones, such as teacher-monitored collective class discussion, transmission of knowledge and individual work, meaning that CL is often in contrast to usual instruction.

Additionally, Shwalb & Shwalb (1995) highlighted how cultural social norms can further challenge the implementation of CL. It has been suggested that for CL to succeed teachers must share a value system congruent with the goals of CL (Smith, Boulton & Cowie, 1994) this could act as a barrier given that the UK education system seems to favour teacher-centred education and independent achievement (Shwalb & Shwalb, 1995). Teachers who believe pupils construct their learning in

social interactions are more likely to use CL in their class (Buch et al., 2017). However, teachers who were educated within a teacher-centred education are still thought to be difficult to convert to the philosophy of CL (Shwalb & Shwalb, 1995).

Jolliffe & Snaith (2017) also indicated that teachers' concerns about behaviour management and time constraints can impact on successful implementation. Sharan (2010) also argued that teachers' perceptions of CL may influence its implementation, suggesting that teachers may view CL as an entity or finished product which they can introduce with immediate success without a clear understanding of the key components for successful implementation. One way of evaluating the implementation of an intervention is to measure fidelity: argued to be just as critical as measuring the intervention itself (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005).

1.5.7.1 Intervention fidelity

Two factors of intervention fidelity have been identified: Adherence and dosage. Dosage refers to how much of the programme was actually delivered and adherence refers to the level of correspondence between the intended intervention and the versions implemented (Dane & Schneider, 1998). The relationship between dosage and outcomes is unclear (Mendive, Weiland, & Snow, 2016) with no a definitive figure as to how long an intervention should be to have a positive effect. Casey, Dyson, & Campbell (2009),who conducted an action research project implementing CL, indicated it may take a few weeks for students to feel comfortable working in a group and to begin cooperating. Within the literature the duration for CL has varied substantially with interventions occurring over varying durations from 3 weeks (e.g. Bertucci, Conte, Johnson, & Johnson, 2010) to 10 weeks (e.g.. Piercy, Wilton, & Townsend, 2002) to 8 months (e.g. Putnam, Markovchick, Johnson, & Johnson, 1996). Nonetheless despite potential challenges in successful implementation, CL is still reported to have positive benefits across a number of academic and social outcomes which will now be discussed.

1.6 Reported benefits of CL

Research on CL has been described as one of the greatest success stories in educational research (Slavin, 1995) with CL advocated as the 'solution for an astonishing array of

educational problems' (Slavin, 1991). There are a number of benefits associated with CL with a dual emphasis on academic and social outcomes with research demonstrating good effects sizes of 0.58 to 0.70 on academic and social variables when compared to control groups (see Johnson & Johnson, 2002). According to Hattie (2009) these effect sizes are notable and could make a real difference in educational interventions. With CL featuring as one of the top classroom methods recommended by Hattie (2009) and the Education Endowment Fund. Although these results are positive, caution does need to be taken when interpreting these findings as they may include small scale studies, which tend to show a larger effect size. Additionally, it cannot be assumed that CL in one study is the same as others, as the models of CL may differ making results less comparable. A further consideration when reviewing the evidence in relation to CL is the cultural and educational contexts within which the research has taken place. CL originates in the USA with research by Johnson and Johnson dominating the research literature therefore it could be questioned as to whether these research findings are applicable to a UK classroom as there may be potential differences in classroom ecology and values. Research in relation to academic and social outcomes will now be discussed.

1.6.1 Academic outcomes

'Hundreds of studies have been conducted examining the effects of CL on productivity and achievement' (Marr, 1997, p10). In an early meta-analysis Johnson, Maruyama, Johnson, Nelson, & Skon (1981) reviewed 122 studies comparing the effectiveness of cooperative, competitive and individualistic goal structures on achievement. Results indicated cooperation is considerably more effective than competition in promoting achievement and productivity. Similar outcomes have been found in comparison to control groups (Slavin, 1996). Similarly, in a meta-analysis Roseth, Johnson & Johnson (2008) reported CL tends to increase student achievement with higher achievement associated with cooperative rather than competitive and individualistic goal structures.

Further evaluation studies have reported pupils' using CL, compared to cooperative or individualistic learning, had a higher recall and reasoning scores (Johnson, Johnson, & Taylor, 1993) and greater retention and higher-level learning (Peterson, Johnson, & Johnson, 1991). CL has also demonstrated positive outcomes in curriculum subjects

such as maths with medium effect sizes (Artut, 2009; Zakaria, Solfiri, Daud, & Abidin, 2013; Capar & Tarim, 2015), Science (Topping et al., 2011a) and Physical Education (Andre, Deneuve, & Louvet, 2011). Overall, students engaging in CL are thought to demonstrate higher order thinking skills, increased retention of information and increased generation of ideas and solutions (Marr, 1997).

1.6.2 Social outcomes

In addition to improvements in student achievement, CL is thought to encourage students to care about one another regardless of differences in race, ability or disability (Johnson, Johnson, & Anderson, 1983). As a result, students' self-esteem, socialisation skills and interpersonal skills are thought to be enhanced (Johnson & Johnson, 1989). People who cooperate are proposed to learn to like one another (Slavin, 1991) as they begin to form supportive personal relationships and a sense of mutual accomplishment (Johnson et al., 1991). It is therefore, not surprising that CL is suggested as a method for promoting social outcomes such as reducing antisocial behaviour and loneliness and improving prosocial behaviours and self-esteem (Johnson et al., 2000).

Pupils in CL conditions, compared to competitive or individualistic learning, have demonstrated higher academic self-esteem and greater cohesion (Johnson, Johnson, & Taylor, 1993). Differences have also been noted in children's social interaction, following CL, with children providing more explanation in response to requests for help and actively listening (Gillies, 2002). For children who have experienced CL, differences have been noted in their social interactions. Gillies (2006) found students in cooperative groups were engaged in more helpful and supportive behaviours to their peers.

It could be questioned, however, whether the successful social outcomes outlined are likely to be dependent on a positive classroom climate which enables pupils to CL tasks together and share ideas without criticism. This is thought to be dependent on pupil interactions and perception of equal status (Jules, 1991). Research indicates that teacher practices can be influential in providing a model for peers to follow, either increasing or decreasing the likelihood of peers becoming more accepted, (Hughes, Cavell, & Willson, 2001) and shaping the peer ecology within the classroom (Hendrickx, Mainhard, Boor-Klip, Cillessen, & Brekelmans, 2015).

1.7 Classroom peer ecology

Classroom peer ecology describes how individuals are nested within social settings, encompassing the richness of interpersonal ties, the status hierarchy and patterns of social behaviours displayed by peers (Hendrickx, Mainhard, Boor-Klip, Cillessen, & Brekelmans, 2015). The notion of an ecology is strongly associated with Bronfenbrenner (1978) and is the part of a child's microsystem with children interacting and influencing one another (Rodkin & Gest, 2010). A richness of interpersonal ties indicates how many positive and how few negative relationships are present (Hendrickx, Mainhard, Boor-Klip, Cillessen, & Brekelmans, 2015). Status hierarchy refers to the structure of social status within the classroom. Status hierarchies can be egalitarian, where each student has a similar status, or hierarchical, where a small group of students have a relatively high peer status. The final aspect of peer ecology is the social behaviours which characterise the interactions among peers.

Peer ecologies characterised by more egalitarian status hierarchies, strong group norms in support of academic achievement, pro-social behaviour, little aggression and little homophily (social clustering based on similarities) are thought to promote positive outcomes (Rodkin & Gest, 2010). The structure and management of the classroom has also been found to influence factors such as bullying (Roland & Galloway, 2002). A more pronounced hierarchy has been linked to increased negative behaviour, specifically victimisation and bullying, compared to more equally distributed classrooms (Schäfer, Korn, Brodbeck, Wolke, & Schulz, 2004).

There seems to be an overlap between some of these characteristics and those promoted by CL, questioning whether CL could play a part in shaping the classroom peer ecology and have a positive impact on three main factors; PA, empathy and bullying. Each of these concepts will now be discussed in turn before drawing them together in relation to CL.

1.8 PA, empathy and bullying

1.8.1 PA

Children's relationships with their peers and friends are associated with multiple aspects of development and adjustment (Gifford-Smith & Brownell, 2003). PA differs

from other aspects of peer functioning, such as friendship, and refers to the degree to which children are liked or disliked by the children in their peer group (Gifford-Smith & Brownell, 2003).

PA has important implications as social rejection by a peer group early in primary schools is associated with later antisocial behaviour (Dodge et al., 2003), reduced classroom participation and increased school avoidance (Buhs, Ladd, & Herald, 2006).

Baumeister & Leary (1995) propose these difficulties reflect emotional and behavioural reactions to perceived threats to social relationships and a failure to have belongingness needs met. A child who is accepted learns cooperation skills, empathy, perspective taking and problem-solving skills with aggressive behaviours tending to decrease (Coie & Dodge, 1998). Whereas low PA has been found to predict increases in relational victimisation (Godleski, Kamper, Ostrov, Hart, & Hall, 2016).

PA also has links to academic achievement with research highlighting correlational findings linking PA with successful academic performance and low levels of acceptance with academic difficulties (e.g. De Rosier, Cillessen, Coie, & Dodge, 1994; Diehl, Lemerise, Caverly, Ramsay, & Roberts, 1998). Once a child becomes rejected there are a number of factors which are thought to maintain this (Coie, 1990), with research highlighting that approximately 50% of rejected children remained rejected over a 5-year period (Coie & Dodge, 1983).

1.8.2 Theoretical frameworks and PA

There are at least two theories which support the view that positive peer relationships and acceptance enhance academic achievement and social development.

1.8.2.1 Hierarchy of needs theory

Maslow's (1970) hierarchy of needs theory states that one of people's basic needs is the need to obtain love and acceptance from family and peers and these needs must be met before growth can occur in other areas. Therefore, from this perspective, if a students' need for acceptance is fulfilled then they should be able to reach achievement goals (Roseth, Johnson, & Johnson, 2008).

1.8.2.2 Belongingness hypothesis

Closely related to Maslow's theory is the belongingness hypothesis. Baumeister & Leary (1995) propose *'Human beings have a pervasive drive to form and maintain at least a minimum quantity of lasting, positive and significant interpersonal relationships' (p 497).* A need to belong is therefore thought to be a fundamental human motivation. Translating this to the school context, the basic premise is that supportive and caring relationships help promote student engagement and adaptive school behaviours (Roseth, Johnson & Johnson, 2008). Similar theoretical support can be drawn to the idea of school connectedness and self-determination theory. Self-determination theory proposes that people's psychological needs are the basis for their self-motivation (Ryan & Deci, 2000). School connectedness describes the quality of social relationships within schools (Rowe, Stewart, & Patterson, 2007). Although there has been no research exploring direct links between CL and connectedness, CL is one teaching approach which has been advocated as building relationships and connectedness (see Rowe et al., 2007).

1.8.3 PA and CL

CL is thought to be the most promising method of instruction for encouraging positive interaction between pupils, particularly between students with Special Educational Needs (SEN) and their peers (Putnam, Markovchick, Johnson, & Johnson, 1996). Over 60 studies (see Johnson et al., 1981; Johnson & Johnson, 2002) have demonstrated the positive effects of using CL in classrooms which include children with SEN. However, most of the research is conducted over a short period, therefore the longevity of this impact on acceptance throughout the school year is not known (Putnam et al., 1996).

Frederickson & Furnham (2004) draw upon social exchange theory to account for differences between children's behavioural characteristics and their peer groups acceptance or rejection. Social exchange theory holds that children will interact with others where the benefits of interaction (in terms of enjoyment, access to resources, achievement of success, feeling good about oneself and receiving praise from adults outweighs the cost (for example in terms of compromise, sharing one's own resources and being the target of undesirable behaviour) (Frederickson & Furnham, 2004). Similarly, Social interdependence theory predicts CL will promote more positive social relationships than individualistic or competitive learning as cooperative goal

structures result in positive cathexis (emotional investment) which resulting in more positive relationships (Roseth, Johnson, & Johnson, 2008).

It could be hypothesised that CL gives children the opportunity to increase their social network in the classroom and provides an opportunity to discuss interpersonal feelings and behaviours and consider other's feelings and points of view. From a peer ecology perspective if an increased number of pupils are accepted this could impact on the number of positive ties within the classroom creating a rich environment and could impact the social hierarchy and subsequent factors. Children who are accepted by their peers are thought to benefit by gaining support and learning empathy (Bierman, 2011). Interestingly, empathy is also thought to play a major role in victimisation experiences, and it could therefore be questioned whether empathy is an important mechanism within PA and peer ecologies.

1.8.4 Empathy

Empathy is thought to be a central building block for successful interpersonal relationships (Reid et al., 2013) and learning (Garton & Gringart, 2005). Having an ability to empathise is considered a prerequisite for regulating behaviour and behaving in a pro-social manner (Belacchi & Farina, 2012). Empathy has been defined as an emotional response or affective state which stems from another's emotional state of condition (Eisenberg & Fabes, 1990). There is a general consensus within the literature that empathy describes the ability to put oneself in the mind of another person (Davis, 1980), however, most definitions of empathy incorporate at least two fundamental elements: affective and cognitive (Reid et al., 2013). The term affective empathy is generally used to refer to having an affective response congruent with that of another's emotional state; and cognitive empathy refers to the ability to understand intellectually the perspective of another person, and in so doing understand another's emotional state (Davis, 1980).

1.8.4.1 Development of empathy

An extension of social interdependence theory (Deutsch, 1949), introduced in paragraph 1.3.1, known as social judgement theory provides one explanation as to how empathy develops (Johnson & Johnson, 2005). Social Judgment theory focuses on relationships between diverse individuals. Contact between diverse individuals results in social judgements that either increase or decrease the liking or hostility they feel

towards each other. Whether hostility or liking results from the contact is thought to be dependent on whether a process of acceptance or rejection takes place. Acceptance is based on the individual's promotion of mutual goal accomplishment because of their perceived positive interdependence. As a result, the psychological processes of substitutability, positive cathexis (emotional investment) and inducibility (openness to being influence and to influencing others (Deutsch, 1949)) are promoted (Johnson & Johnson, 2005). It is suggested that when positive cathexis and inducibility occurs, liking for the person engaging in effective actions results (Johnson & Johnson, 2005) leading to promotive interaction which creates frequent and open communication and an understanding of each other's perspectives which leads to empathy (Johnson & Johnson, 2005).

Empathy has been linked to social functioning and prosocial behaviour in pre-school children (Williams, O'Driscoll, & Moore, 2014), primary aged (Roberts, Strayer, & Denham, 2014) and high school-aged children (Paciello, Fida, Cerniglia, Tramontano, & Cole, 2013). Empathy has also been linked to victimisation, with Malti, Perren, & Buchmann (2010) finding that increases in victimisation predicted decreases in empathy. Early research exploring empathy directly has shown that training in empathy can lead to reductions in aggression and increases in pro social behaviour and self-esteem (Feshbach, 1984). These findings highlight the potential for empathy to act as a mediating variable in intergroup relations with empathy being incorporated into group relation interventions (Stephen & Finlay, 1999). The link between empathy and CL, specifically, will now be discussed.

1.8.5 Empathy and CL

CL is thought to help develop empathy and cohesiveness creating a classroom environment where the practice of responsibility, fairness, tolerance, teamwork and understanding and respect for different points of views is a central feature (Baker, 2004). Research suggests that working in heterogenous groups and collaborating with peers results in increased empathy, altruism and an ability to view situations from a variety of perspectives (Johnson & Johnson, 1992).

Although research exploring the link between CL and empathy directly is limited, Bridgeman (1981) found that the role taking of children in cooperative jigsaw groups increased over the course of the 8 weeks in contrast to scores in control classes.

Goudas & Magotsiou (2009) examined the effects of a cooperative physical education program and found that students who participated in the programme increased their cooperative skills and empathy and decreased their tendency to disrupt. These results could be explained by the proposed relationships between empathy and prosocial behaviour, as empathy is thought to facilitate prosocial behaviour and inhibit anti-social behaviour (Jolliffe & Farrington, 2006) with prosocial children found to show greater empathic awareness (Warden & Mackinnon, 2003).

High empathy is proposed to facilitate actions to benefit others whereas those with low empathy may fail to connect their anti-social behaviour to the emotional reactions of others (Jolliffe & Farrington, 2006). This has led to suggestions within the literature that low empathy and bullying maybe be causally related based on a theoretical framework that those who engage in bullying have less empathy than those who do not (Jolliffe & Farrington, 2006).

1.9 Bullying

A definition of bullying was developed by Olweus (1993) who defined bullying as being 'exposed, repeatedly and over time, to negative actions on the part of one or more other students' (p.9). Bullying has also been described as aggressive behaviour or intentional harm that is carried out repeatedly and characterised by a perceived imbalance of power or strength (Dan Olweus & Limber, 2010). The term peer victimisation is sometimes used within the literature to describe similar experiences.

Theory and research into the social aspects of bullying and victimisation play an important role in understanding this behaviour. A phenomenon, known as homophily, explains that individuals are likely to select friends based upon similarities in behaviour and/or attitudes (see Brechwald & Prinstein, 2011; Dishion, Patterson, Stoolmiller, & Skinner, 1991). Additionally, within groups of bullies, social norms are established which view bullying positively and these norms can elicit and reinforce bullying behaviour among group members (Hong & Espelage, 2012).

A widespread approach to preventing victimisation is the Olweus (1993) programme, which covers topics related to bullying (described as harassment or aggressive behaviour directed at another in the context of a power balance) and victimisation, focussing on awareness, empathy and attitudes. However, no strong empirical

justification for the approach has been established and results have not been universally consistent across contexts (Olweus & Limber, 2010).

Previous reviews have concluded that anti-bullying programs have little effect (e.g. Merrell, Gueldner, Ross, & Isava, 2008). Although, Ttofi & Farrington (2011) conducted a systematic review and found decreases in bullying and victimisation following anti-bullying programs and a handful of studies have been shown to address bullying and victimisation with positive outcomes (Frey, Hirschstein, Edstrom, & Snell, 2009), with whole school approaches most effective (e.g. Vreeman & Carroll, 2007).

Bullying has traditionally been associated with social dominance theories. Social dominance is defined as 'a differential ability to control resources using coercive and prosocial strategies' (Hawley, 1999; p.105). The importance of a dominance hierarchy in a group is based on the idea that competition for scarce social and material resources underlies all human relations. Social dominance theories propose that a dominance hierarchy exists within every social situation based on competition for resources, which are obtained through bullying or through the strategic use of prosocial behaviours (Hawley, 1999). This idea gives rise to the hypothesis that the use of pro social behaviours needs to be promoted to avoid children engaging in bullying behaviour. However, children's social behaviours are more complex. In educational settings prosocial behaviour is not always valued and children who behave in 'prosocial ways' by sharing and helping others may not be popular with their peers (Warden & Mackinnon, 2003). Research into peer ecology would seems to indicate that to promote pro-social behaviours a change needs to occur within the ecology of the classroom. It is thought that the more cooperative experiences children have the more likely it is they will engage prosocial behaviours (Choi, Johnson, & Johnson, 2011) and therefore it could be suggested are less likely to engage in bullying behaviours.

1.9.1 Bullying and CL

In cooperative contexts, it is expected students will develop a genuine sense of care and respect for their peers and as differences in power are reduced, opportunities for bullying are also thought to reduce (Hymel et al., 2015). Consistent with this argument, research has demonstrated that as children's exposure to CL increased prosocial

behaviour also increases (Choi, Johnson & Johnson, 2011). Students who were briefly exposed to a cooperative, neutral or competitive norm in game situations expressed more regret about hypothetical bullying when exposed to cooperative norms (Choi, Johnson, & Johnson, 2011). Additionally, in a meta-analytic review of 44 anti-bullying programs in school, co-operative group work was one of the most important elements associated with a decrease in bullying (Ttofi & Farrington, 2011).

Social interdependence theory suggests there is a dominance hierarchy only in competitive situations (not cooperative or individualistic) (Choi et al., 2011). In cooperative situations, individuals obtain resources through prosocial actions to promote the success of groupmates, whereas in an individualistic situation individuals obtain resources for their own benefit only. It is postulated that in cooperative situations individuals would avoid seeking dominance, whereas in individualistic or competitive situations individuals may be bullies or victims (Choi et al., 2011). Social interdependence theory would, therefore, argue that in CL contexts individuals are more likely to be equal whereas in individualistic situations individuals are more likely to see themselves as separate and are therefore more likely to be bullies or victims (Choi, Johnson, & Johnson, 2011).

However, putting students in a group does not guarantee positive interactions will occur (Van Ryzin & Roseth, 2018). Social psychology research demonstrates through an understanding of stereotyping, prejudice and discrimination that trying to facilitate contact between students who belong to different groups may exacerbate differences (Dovido, Gaertner, & Saguy, 2009). It is argued that to promote social integration, the breakdown of biases and prejudices among students who belong to different groups must be promoted (Pettigrew, 1998). Positive interdependence is suggested as the key ingredient for promoting this interaction (Johnson et al., 1981). Deutsch (1949) proposes that in situations where positive interdependence is promoted peer interactions change and instead of competing with or ignoring one another peers are more likely to promote success. It is argued that in turn these positive interactions increase interpersonal attraction and acceptance whilst reducing social isolation and supporting the development of new friendships (Johnson, Johnson, Roseth, & Shin, 2014; Mikami, Boucher, & Humphreys, 2005).

Choi, Johnson, & Johnson's (2011) research has highlighted a positive association between cooperativeness and prosocial behaviours and a positive relationship between competitiveness and physical, relational and verbal aggression. This supports the idea that perceived positive interdependence leads to promotive interaction whereas perceived negative interdependence leads to oppositional interaction (Roseth, Johnson & Johnson, 2008). Choi, Johnson and Johnson (2000) argue the results demonstrate that social interdependence theory can be a helpful way to examine dominance and power in the classroom.

Despite clear theoretical links which could be argued to underpin CL and bullying, to the author's knowledge limited research has been undertaken in these areas, although this is starting to be explored. For example, Van Ryzin & Roseth (2018) used a CL approach based on Johnson's conceptual model and explored the impact on victimisation and bullying. Van Ryzin & Roseth (2018) have started to pave the way for research to explore whether an intervention focussing solely on increasing cooperative activities can, through intervening with the peer ecology, reduce bullying and victimisation. However, like other CL research, this is conducted outside of the UK and uses a CL approach which may be difficult to implement without extensive training.

1.10 A joining of cooperative learning, empathy, PA and bullying

The literature so far has highlighted the importance of PA for a range of academic, social and psychological outcomes. In CL groups work together to meet common learning goals, requiring positive interdependence and resulting in promotive interaction as individuals encourage each other to reach the group's goals (Johnson & Johnson, 1992). This cooperation with peers is thought to result in increased empathy, altruism and perspective taking (Johnson & Johnson, 1992) and PA (Craig, 2010; Putnam et al., 1996; Smith, 2017).

There are similarities which can be drawn between CL and the development of PA, empathy and the prevention of bullying. Social interdependence theory and the promotion of positive interdependence is a key theoretical framework underpinning these concepts. Bullying has been associated with low empathy (Endresen & Olweus, 2001) and low PA (de Bruyn, Cillessen, & Wissink, 2010) with PA thought to act as a

protective factor against bullying and victimisation (Pellegrini & Bartini, 2000). As a result, in CL contexts it is expected that as students develop a genuine sense of care and respect for their peers and power differences reduce, opportunities for bullying would reduce (Hymel et al., 2015).

In order to gain a greater understanding of the evidence base regarding the impact of CL on PA, empathy and bullying, a systematic literature review was undertaken.

1.11 Systematic literature review

The aim of the review was to systematically identify research exploring the impact of CL on PA, empathy and bullying behaviour. There is not standard process for conducting systematic reviews with authors suggesting different stages to conducting a review (e.g. Perestelo-Perez, 2013). Gough (2007) outlines a series of stages for systematic review which incorporates many of the stages seen in other methods (see figure 1.1). These stages were followed throughout the systematic review process to ensure the process and methods were as transparent as possible:

Stages of Review (Gough, 2007)

- Stage 1: Formulate review question and develop a protocol
- Stage 2: Define studies to be considered (Inclusion Criteria)
- Stage 3: Search for Studies (Search Strategy)
- Stage 3: Screen Studies (check that they meet inclusion criteria)
- Stage 4: Describe studies (Systematic map of research)

1.11.1 Review Question

A research question needed to be clearly defined. This was done using the PICO method (Petticrew & Roberts, 2006), which defines the population, intervention, comparison and outcomes. The following research question was developed to be able to review an appropriate body of literature:

What is the impact of CL on pupils' PA, empathy and bullying behaviour?

The PICO approach provides a breakdown of the research question:

Population: Children and young people aged between 4 and 16 years old.

Intervention: Cooperative Learning

Comparison: Control group or alternative intervention

Outcome: Pre and post measures to assess whether the intervention has had an effect on PA, empathy or bullying.

1.11.2 Criteria for considering studies for the review (Inclusion/Exclusion criteria).

The following inclusion/exclusion criteria were used to establish the most relevant studies to answer the review question.

1. The study was published in a peer review paper. Unpublished research or review articles were excluded from the review.

2. The full paper was available in English. Studies which were not written in English were excluded from the review.

4. The study evaluated the impact of a CL intervention regarding one of the relevant outcomes. Studies which did not include one of these outcomes were excluded.

5. The study employed an experimental design, with repeated measures. Studies which were descriptive or had no quantitative data were excluded as they do not provide evidence as to the effectiveness of interventions.

6. The intervention took place within a school setting with participants aged between 4 and 16 years old.

7. The full article was available

1.11.3 The search strategy

An initial search for studies relevant to the research question was carried out on 6th July 2018. A final search for studies was carried out on 21st May 2019 to take into consideration any recent publications. The following electronic databases were searched:

- PsycINFO
- EBSCO Host

- NU Search

The following search terms were used:

‘Cooperative Learning’ AND ‘children’

‘Cooperative Learning’ AND ‘children’ AND ‘PA’

‘Cooperative Learning’ AND ‘children’ AND ‘Empathy’

‘Cooperative Learning’ AND ‘children’ AND ‘bullying’ OR ‘victimisation/Victimization or ‘peer relations’

‘Cooperative Learning’ AND ‘children’ AND ‘PA’ AND ‘Empathy’ AND ‘bullying’

In order to establish the scope of the literature a search using the search terms ‘Cooperative Learning’ AND ‘children’ AND PA’ AND ‘Empathy’ AND ‘bullying’ was employed first, however as this produced no results, the decision was taken to separate the search terms and start with a general search term of ‘Cooperative Learning AND ‘children’. This produced 101 results and following this additional search terms related to the research i.e. PA, empathy and bullying were added.

Where the database allowed limits were put in place to only include peer-reviewed journals and those written in English in accordance with the inclusion criteria.

1.11.4 Studies included in the systematic literature review

The initial search returned 142 records. Following a screening of the titles and abstract and application of the inclusion criteria 105 records were excluded in line with inclusion criteria. 37 articles remained from across all searches. Of these papers, 15 were excluded because they were duplicate articles or were not written in English. A full text of 22 papers was assessed for further eligibility. 14 papers which were felt to meet the inclusion criteria were obtained. Of these 6 papers were excluded for reasons such as the approach was not considered to be synonymous with CL, an experimental methodology was not used, or dependent variables were not considered to be closely related to one of the three variables set. The eight remaining studies were identified for the systematic review. A flow chart of the search process can be found in Figure 1.1. A summary of search results can be found in Appendix A and references of studies

excluded at full text can be found in Appendix B. Table. 1.1 provides a summary of the key information extracted from the eight studies included in the review

Figure 1. 1 A flow diagram summarises the search process

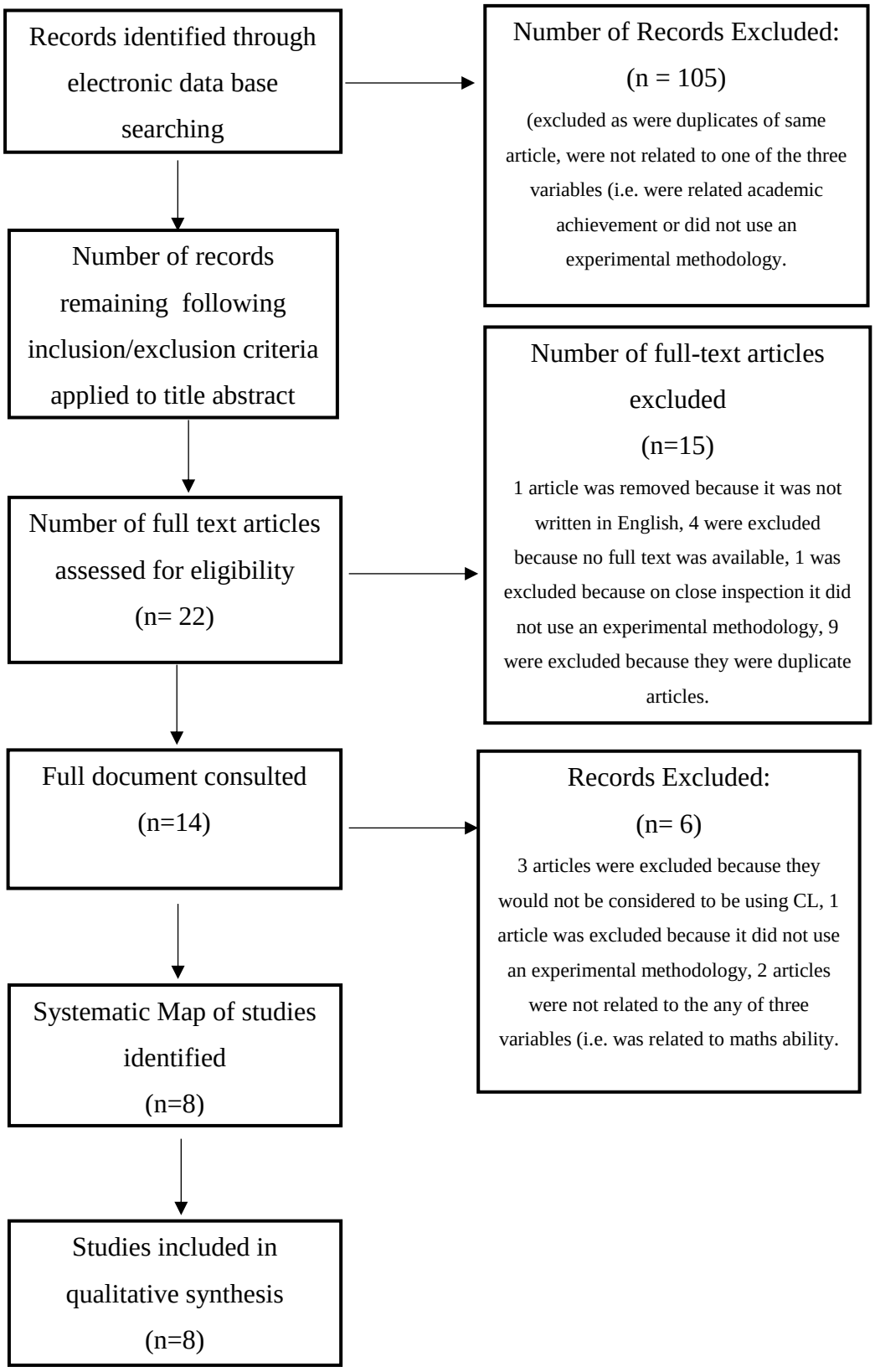


Table 1. 1 Summary of key information extracted from studies

Study	Publication and Location	Outcome Measure	Sample	Design	Intervention		Analysis and Findings
Capodieci, Rivetti, & Cornoldi (2016)	Journal of Attention Disorders Italy	PA	30 children with ADHD (aged between 6 and 10years)	Random allocation to either: 1) CL (Experimental) 2) Standard practices (Control)	**Focus of CL**	Usual Curriculum	2x2 ANOVA Improvements in the socio-metric status of children with ADHD in CL classes (n^2_p =.14) Reduction and improvement in inattention (n^2_p =.31) hyperactivity (n^2_p =.42) social ability (n^2_p =.29) of ADHD in CL and control group.
					Training and Facilitators	Facilitated by teachers following staff training	
					Model of CL	Manual for personalised CL (Rivetti & Capodieci)	
					Social Skills	Not specified	
					Frequency and Duration	6 sessions, one per week, 2-hour duration (approx.. exposure =12 hours)	
					Intervention Fidelity	Teacher observations and meeting with teachers of both groups	

Jacques, Wilton, & Townsend (1998)	Journal of Intellectual Disability Research New Zealand	PA	24 pupils (12(m), 3(f)) aged 9-11 with 'mild intellectual disability'	Random allocation to either : 1) CL (experimental) 2) usual classroom (control) 5 week follow up Pre/post measures	**Focus of CL**	Geography and Social Studies	Three way analyses of varaiance. A significant increase in social acceptance of the focus pupil in CL classes ($p<.001$) but not in control classes. Effects maintained at a 5 week follow up. CL was more effective for focus pupil in mainstream class compared to special needs classes.
					Training and Facilitators	Researcher-Led	
					Model of CL	Slavin (1983) Jigsaw activities	
					Social Skills	Not specified	
					Frequency and Duration	6 weeks (4x30 minute weekly sessions. Approx. exposure =12 hours)	
					Intervention Fidelity	Not specified	

							No effect sizes reported
Piercy, Wilton, & Townsend (2002)	American Journal on Mental Retardation New Zealand	PA	51 pupils (29m and 22f) aged 5-7 including 6 pupils with 'moderate to severe intellectual disability'	Random allocation to either: 1) CL (experimental) (n=17) 2) Social Contact (comparison) (n=17) 3) Usual practices (control) (n=17)	**Focus of CL**	Johnson and Johnson (1999) model of CL	3x2 repeared measures ANOVA A significant increase for PA and popularity ratings for focus pupil in CL condition ($p<.004$). No effect sizes reports No significant change in the control for comparison condition (social contact)
					Training and Facilitators	Teacher-led – training details not specified	
					Model of CL	Johnson and Johnson (1999) CL model	
					Social Skills	Not specified	
					Frequency and Duration	10 weeks (2x40 minute weekly sessions)	
					Intervention Fidelity	Not specified	
Putnam et al (1996)		PA	214 pupils (aged 10-15)	Quasi-experimental	**Focus of CL**	Usual Curriculum	ANOVA

	The Journal of Social Psychology USA		across 2 schools. 41 pupils were described as 'disabled'	1) Experimental – CL 2) Comparison - Competitive learning	**Training and Facilitators**	Teachers received 34 hours teaching training over 6 months	Significant increases in pre to post test for PA in CL condition ($p<.001$), no significant change in control condition. PA in control conditions was stable across the study period
					Model of CL	Johnson and Johnson 5 principles	
					Social Skills	No social skills training	
					Frequency and Duration	8 months (2x45 minutes lessons weekly, approx. exposure = 42-48 hours)	
					Intervention Fidelity	Not specified	
Oortwijn, Boekaerts, Vedder & Strijbos (2008)	International Journal of Educational Psychology	PA	94 pupils 10-12 years from 5 multi-ethnic elementary schools	Quasi-experimental	**Focus of CL**	Maths curriculum	Learning gains were predicated by peer tutoring.
					Training and Facilitators	Teacher facilitated with two-hour mini training workshop	

	The Netherlands				**Model of CL**	Not specified	
					Social Skills	2x1 hour social skills training	
					Frequency and Duration	11x 1-hour sessions	
					Intervention Fidelity	Not specified	
Quinn (2002)	Education and Treatment of Children USA	Anti-social behaviour	15 first grade students (n=15). Students selected based on teacher nominations of students who showed antisocial behavioural characteristics.	Random allocation to either: Condition A (Treatment condition) Condition B (Control condition)	**Focus of CL**	CL lessons outside of the usual curriculum	A significant in academic engaged time, however no significant differences in externalising antisocial behaviours or negative playground interactions.
					Training and Facilitators	Teacher facilitated	
					Model of CL	Stahl (1994)	
					Social Skills	5 sessions focussing on active communication skills	
					Frequency and Duration	5 weeks (26 x20 minutes daily)	

					Intervention Fidelity	Meeting with teacher 2 x week, detailed lesson plans provided and random observations.	
Van Ryzin & Roseth (2018)	Journal of Educational Psychology USA	Bullying and Victimisation	1460 7th grade students from 15 middle schools	Cluster randomised control trial with pre-post measures. Intervention (7 schools) and Control Condition (8 schools)	**Focus of CL**	Usual Curriculum	CL significantly reduced bullying, victimisation and perceived stress for marginalised students and reduced emotional problems and enhanced relatedness for all students. No effect sizes reported.
					Training and Facilitators	Teacher facilitated with three half-day training sessions	
					Model of CL	Johnson and Johnson conceptual approach	
					Social Skills	Not specified	
					Frequency and Duration	Not specified	
					Intervention Fidelity	Teacher Observations	

Vasileiadou (2009)	Emotional and Behavioural Difficulties Cyprus	PA	Kindergarten (pre-primary) 17 children (6 m, 11 f) were focus	Pre and Post Measures	**Focus of CL**	Specially designed activities	9 out of 17 children had an increase in the number of votes they received. No inferential analysis reported.
					Training and Facilitators	Researcher-led	
					Model of CL	Johnson and Johnson 'learning together' method	
					Social Skills	1-week social skills	
					Frequency and Duration	6 weeks with 3-week intervention weeks with 30-minute session daily	
					Intervention Fidelity	Not specified	

1.11.5 Appraisal of study quality and relevance

The quality and relevance of each study were assessed using a rating-based system as this is seen as a crucial part of the review process (Gough, Oliver & Thomas, 2013). Randomised Controlled Trials (RCTs) could be considered most effective for the research question (Gough et al, 2013). However, limiting the criteria to only include RCTs may not provide good evidence of what works in real educational settings and other quality designs may be omitted.

Gough (2007) outlines a framework in which a research study is assessed against criteria to determine the 'weight of evidence' of the findings in being able to answer the research question. This allows for the inclusion of all designs but places less emphasis on designs with a lower rating. The weight of evidence model uses three different judgement criteria: A, B and C. Each criterion are described in Table 1.2.

Table 1. 2 Outline of the Weight of Evidence Criteria used within the research

Criteria	**Description**
WoE A: Quality of the methodology	The coherence and integrity of the evidence in its own terms
WoE B: Methodological Appropriateness	The relevance of the methodology for answering the review question
WoE C: Relevance of Evidence	The relevance of the study for answering the review question
WoE D: Overall Weight of Evidence	Overall judgement of the study based upon the ratings given for A, B and C

The development of the WoE criteria was informed by Gersten et al. (2005) who identified quality indicators for quasi-experimental research and by Smith (2018) who developed WoE criteria for a study exploring CL. A copy of the criteria used to determine the WoE can be found in Appendix C.

An analysis of the eight included studies was undertaken to determine each papers WoE. A copy of this analysis can be found in table 1.3. A summary of the titles of the eight studies included in the review can be found in table 1.4.

Table 1. 3 Weight of Evidence analysis for the studies included in the systemic review

WoE A Criteria	**Capodieci, Rivetti & Conoldi (2016)**	**Jacques, Wilton & Townsend (1998)**	**Oortwijn, Boekaerts, Vedder & Fortuin (2008)**	**Piercy, Wilton & Townsend (2002)**	**Putnam et al (1996)**	**Quinn (2002)**	**Ryzin & Roseth (2018)**	**Vasileiad-ou (2009)**
1. Hypotheses /aims/ objectives are clearly described	✓	✓	✓	✓	✓	✓	✓	✓
2. Main outcomes to be measures are clearly described in the introduction or methods section	✓	✓	✓	✓	✓	✓	✓	✓
3. Participants characteristics are clearly described (e.g. number, age, gender).	✓	✓	✓	✓	✓	✓	✓	✓
4. The Cooperative learning intervention is clearly described (including duration, frequency, facilitator and approach/model used for Cooperative Learning (e.g. Kagan etc).	✓	✓	×	✓	✓	✓	×	×
5. The fidelity of the intervention is clearly described	✓	×	✓	×	✓	✓	✓	×
6. Random assignment to condition	×	✓	×	✓	×	✓	✓	✓
7. Methods and procedures are clearly described	✓	✓	✓	✓	✓	✓	✓	✓
8. Control group and/or comparison group	✓	✓	×	✓	✓	✓	✓	×
9. Reliability/Validity of relevant measures described	✓	✓	✓	×	×	×	✓	×
10. Main findings are described	✓	✓	✓	✓	✓	✓	✓	✓
11. Exact p levels are reported	✓	✓	×	✓	✓	✓	✓	×
12. Limitations are discussed	✓	✓	✓	✓	✓	✓	✓	×
Total rating	**11 (high)**	**11 (high)**	**8 (Medium)**	**10 (high)**	**10 (High)**	**11 (high)**	**11 (high)**	**6 (low)**

WoE B	**Capodieci, Rivetti & Conoldi (2016)**	**Jacques, Wilton & Townsend (1998)**	**Oortwijn, Boekaerts, Vedder & Fortuin (2008)**	**Piercy, Wilton & Townsend (2002)**	**Putnam et al (1996)**	**Quinn (2002)**	**Ryzin & Roseth (2018)**	**Vasileiadou (2009)**
1. Randomised Control Trial (RCT) or quasi-experimental design	✓	✓	✓	✓	✓	✓	✓	✓
2. Random allocation to groups	✓	✓	×	✓	×	✓	✓	×
3. Participants appropriate (e.g. age, setting)	✓	✓	✓	✓	✓	✓	×	×
4. Control or comparison group	✓	✓	×	✓	✓	✓	✓	×
5. Measures used are related to the research question	✓	✓	✓	✓	✓	✓	✓	✓
6. Pre and post data collected	✓	✓	✓	✓	✓	✓	✓	✓
7. Social Skills Training is present	✓	×	×	×	×	✓	×	×
8. Teacher selected Heterogenous groups	✓	✓	✓	✓	×	✓	×	✓
9. Method of CL is clearly stated	✓	✓	×	✓	×	✓	✓	✓
Total Score and WoE B rating	**9 (High)**	**8 (High)**	**5 (Medium)**	**8 (High)**	**5 (Medium)**	**9 (High)**	**6 (High)**	**5 (Medium)**

WoE C	Capodieci, Rivetti & Conoldi (2016)	Jacques, Wilton & Townsend (1998)	Oortwijn, Boekaerts, Vedder & Fortuin (2008)	Piercy, Wilton & Townsend (2002)	Putnam et al (1996)	Quinn (2002)	Ryzin & Roseth (2018)	Vasileiadou (2009)
1. The study was based in the UK	×	×	×	×	×	×	×	×
2. Cooperative learning as a whole class intervention	✓	✓	✓	×	✓	✓	✓	×
3. CL delivered by class teachers	✓	×	✓	×	✓	✓	✓	×
4. Participants are predominately in KS2	✓	✓	×	×	×	✓	×	×
5. Participants are in a mainstream setting	✓	✓	✓	✓	✓	✓	✓	✓
6. CL used within an academic lesson and within the usual curriculum	✓	✓	✓	×	✓	✓	✓	×
7. The study evaluates the impact of CL on PA or related construct	✓	✓	✓	✓	✓	×	×	✓
7. The study evaluates the impact of CL on empathy or related construct	×	×	×	×	×	×	×	×
7. The study evaluates the impact of CL on bullying or related construct	×	×	×	×	×	✓	✓	×
8.. Usual teaching as a control group	✓	✓	✓	×	✓	✓	✓	×
9. Participants are in their usual classes	✓	×	✓	✓	✓	✓	✓	×
Total Score Rating WoE C rating	**8 (High)**	**6 (High)**	**6 (High)**	**3 (Low)**	**7 (High)**	**8 (High)**	**7 (High)**	**2 (low)**
WoE D rating	**High**	**High**	**Medium**	**High**	**High**	**High**	**High**	**Low**

Table 1. 4 Author, Title and weight of evidence for studies included in the systematic review

Author	**Title**	**Overall Weight of Evidence**
Capodieci, Rivetti & Conoldi (2016)	A Cooperative Learning Classroom Intervention for Increasing Peer's Acceptance of Children With ADHD	High
Jacques, Wilton & Townsend (1998)	Cooperative learning and social acceptance of children with mild intellectual disability	High
Oortwijn, Boekaerts, Vedder & Fortuin (2008)	The impact of a cooperative learning experience on pupils' popularity, non-cooperativeness, and interethnic bias in multi-ethnic elementary schools	Medium
Piercy, Wilton & Townsend (2002)	Promoting the Social Acceptance of Young Children with Moderate-Severe Intellectual Disabilities Using Cooperative-Learning Techniques	High
Putnam et al (1996)	Cooperative Learning and PA of Students with Learning Disabilities Article	High
Quinn (2002)	Changing Antisocial Behavior Patterns in Young Boys: A Structured Cooperative Learning Approach.	High
Ryzin & Roseth (2018)	Cooperative Learning in Middle School: A Means to Improve Peer Relations and Reduce Victimization, Bullying, and Related Outcomes	High
Vasileiadou (2009)	Cooperative learning and its effects on pre-primary, marginalized children.	Low

1.11.6 Synthesis of findings

A narrative synthesis of the eight studies included in the review is presented. Each of the studies were compared on the sample, research design, intervention, measures and dependent variables and relevant findings. The synthesis aims to provide a clearer picture as to the overall evidence. As only one of the studies looked at CL and bullying this will be summarised at the end of each subsection.

1.11.7 Sample

1.11.7.1 PA

None of the studies were carried out in the UK. Two studies were carried out in New Zealand (Jacques, Wilton, & Townsend, 1998); (Piercy et al., 2002) and two in the USA (Putnam, Markovchick, Johnson, & Johnson, 1996; Quinn, 2002). The other studies were carried out in Italy (Capodieci et al., 2016), The Netherlands (Oortwijn, Boekaerts, Vedder, & Fortuin, 2008) and Cyprus (Vasileiadou, 2009). The participants were from a variety of schools, including pre-primary/kindergarten (Vasileiadou, 2009), primary (Capodieci et al., 2016; Jacques et al., 1998; Piercy et al., 2002; Quinn, 2002), middle and secondary schools (Michiel Bastiaan Oortwijn et al., 2008; Putnam et al., 1996; Van Ryzin & Roseth, 2018). Four studies were run in multiple schools (Capodieci et al., 2016; Oortwijn et al., 2008; Putnam et al., 1996; Van Ryzin & Roseth, 2018).

The age of the participants ranged from 5 years old (Piercy et al., 2002; Vasileiadou, 2009) to 15 years old (Putnam et al., 1996c). Two studies had pupils aged between 6 and 9 (Capodieci, Rivetti, & Cornoldi, 2016; Quinn, 2002), one study had pupils aged between 9 and 11 (Jacques et al., 1998b) and three studies had pupils aged between 10 and 15 (Michiel Bastiaan Oortwijn et al., 2008; Putnam et al., 1996c; Van Ryzin & Roseth, 2018). The number of participants ranged from 13 (Vasileiadou, 2009) to 1460 (Van Ryzin & Roseth, 2018). Two studies involved over 200 children (Putnam et al., 1996; Quinn, 2002) with the remaining five studies involving between 24 and 94 pupils (Capodieci et al., 2016; Jacques et al., 1998; Oortwijn et al., 2008; Piercy et al., 2002). Most studies had an even balance between male and female participants except three studies who had entirely male (Piercy et al., 2002) or predominately male participants (Jacques et al., 1998b).

Bullying

The one study related to bullying was carried out in the USA (Van Ryzin & Roseth, 2018). The study ran across 15 middle schools with 1460 participants, aged 12-13 years with an approximate balance between male and females.

1.11.8 Research design

1.11.8.1 PA

All studies used pre and post measures with either random allocation to condition or quasi-experimental design. One study included a follow up at 5 weeks (Jacques et al., 1998). Five of the eight studies had control or comparison groups (Capodieci et al., 2016; Jacques et al., 1998; Putnam et al., 1996; Quinn, 2002; Van Ryzin & Roseth, 2018), with one study having both a control and a social contact' comparison group (Piercy et al., 2002). In two of the studies all pupils received CL and there was no control or comparison group (Oortwijn et al., 2008; Vasileiadou, 2009).

In five of the studies pupils were taught in their usual classes (Capodieci et al., 2016; Oortwijn, Boekaerts, Vedder, & Strijbos, 2008; Putnam et al., 1996; Quinn, 2002; Van Ryzin & Roseth, 2018). In one study pupils who were considered to have an 'intellectual disability' were placed in different classes for the research (Jacques et al., 1998) and in another study mainstream children attended special education classes to carry out group work (Piercy et al., 2002). In another study, pupils were enrolled into one of two classes (Vasileiadou, 2009).

1.11.8.2 Bullying

In the study related to bullying (Van Ryzin & Roseth, 2018) a cluster randomised control trial design was employed using pre and post measures. The study used random allocation to an intervention or control condition and pupils were taught in their usual classes.

1.11.9 Intervention

Different aspects of the intervention were focussed on as outlined below and each of these studies will be synthesised in relation to each of these elements:

- The focus of CL

- Training and facilitators of the intervention
- Model of CL or approach used
- Social Skills Training
- Frequency and duration of the intervention
- Intervention fidelity

1.11.9.1 The focus of CL

In four of the PA studies the CL intervention was undertaken within the usual curriculum (Agnese Capodieci et al., 2016; Jacques et al., 1998; Oortwijn et al., 2008; Putnam et al., 1996; Van Ryzin & Roseth, 2018) with two studies using CL within specific subject areas such as geography and social studies (Jacques et al., 1998b) and maths (Oortwijn et al., 2008). In the other three studies, activities were specifically designed to be used during the CL intervention which included activities such as drawing and painting (Maureen Piercy, Wilton, & Townsend, 2002; Vasileiadou, 2009). In one study activities were centred around teaching skills in communication, anger management and cognitive and behavioural skills (Quinn, 2002). In the bullying study (Van Ryzin & Roseth, 2018) the CL intervention was used within the usual curriculum

1.11.9.2 Facilitators and training

In most of the PA studies the CL intervention was delivered by the class teachers (Capodieci et al., 2016; Oortwijn et al., 2008; Piercy et al., 2002; Putnam et al., 1996) except in two studies (Jacques et al., 1998b; Vasileiadou, 2009) where it was facilitated by the researcher and staff did not receive any training. The amount of training that staff received varied greatly or was not specified. Teacher training ranged from a 2-hour mini workshop (Oortwijn et al., 2008) to 34 hours of training over a 6 month period (Putnam et al., 1996).

In the studies which looked at bullying, teachers facilitated the CL intervention (Quinn, 2002; Van Ryzin & Roseth, 2018). However, the training varied as one study provided three in-person training session over a period of 12 months (Van Ryzin & Roseth, 2018), whilst the other study did not specify how much training teachers received (Quinn, 2002).

1.11.9.3 CL intervention

Three of the PA studies were based on Johnson and Johnsons conceptual framework which is centred around five principles (Piercy et al., 2002; Putnam et al., 1996; Vasileiadou, 2009). Jacques et al (1998) drew upon Slavin's STAD and Jigsaw activities. One study based the intervention around a manual for personalised cooperative learning by Rivetti & Capodieci (Capodieci et al., 2016). The remaining study did not specify the CL approach taken (Oortwijn et al., 2008).

Out of the two bullying studies one used Johnson and Johnson's conceptual framework, with Johnson and Johnson providing the training for the teachers (Van Ryzin & Roseth, 2018) and one used Stahl's (1994) structured cooperative learning (Quinn, 2002).

1.11.9.4 Social skills training

Out of the seven PA studies, only one made reference to social skills training in some form, which consisted of two sessions at the start of the intervention about rules and helping behaviour (Oortwijn et al., 2008).

Out of the bullying studies, one study made reference to social skills training, which consisted of 5 sessions focussing on teaching active communication skills (Quinn, 2002).

1.11.9.5 Frequency and duration

The duration of the intervention in the PA studies varied from 3 weeks (Vasileiadou, 2009) up to 10 weeks (Piercy et al., 2002) and eight months (Putnam et al., 1996). The remaining studies lasted between 5 and 6 weeks (Capodieci et al., 2016; Jacques et al., 1998;). However, one study did not specify the duration of the intervention (Oortwijn et al., 2008). In terms of frequency studies varied from daily (Jacques et al., 1998b; Vasileiadou, 2009) to twice a week (Piercy et al., 2002; Putnam et al., 1996) to weekly (Capodieci et al., 2016). One study did not specify the frequency of the intervention (Oortwijn et al., 2008). In terms of session duration this varied from 30 minutes (Jacques et al., 1998; Vasileiadou, 2009) to 40 minutes to one hour (Oortwijn et al., 2008; Piercy et al., 2002; Putnam et al., 1996) to two hours (Capodieci et al., 2016).

Of the studies which provided details of the frequency and duration, it was possible to calculate an approximate estimate as to the amount of exposure the participants had to CL. In three PA studies (Capodieci et al., 2016; Jacques, Wilton, & Townsend, 1998; Oortwijn et al., 2008; Piercy et al., 2002) pupils had between 11 and 13 hours of exposure to CL during the intervention period. Two studies had less than this with pupils having approximately 8 hours during the intervention period (Vasileiadou, 2009). During one PA study (Putnam et al., 1996) students had significantly more exposure to CL with students receiving a total of 51 hours over the intervention period which lasted 8 months.

Out of the two bullying studies, one study ran daily sessions which accounted to approximately 8 hours over the intervention period (Quinn, 2002). The other study provided no information in terms of the frequency and duration of the intervention (Van Ryzin & Roseth, 2018).

1.11.9.6 Intervention fidelity

Procedures to ensure the fidelity of the intervention were outlined in four of the PA studies. Two studies included observations as part of a fidelity check to verify CL was being used appropriately in the CL condition (Oortwijn et al., 2008; Putnam et al., 1996). One study selected teachers at random to videotape twice a week and used a coding method to ensure the CL intervention was being adhered to (Oortwijn et al., 2008). One study offered feedback before and after activities and ensured fidelity through meeting the teachers of both the experimental and control group (Capodieci et al., 2016).

In the bullying studies, one study used teacher observations (Van Ryzin & Roseth, 2018) and one used a collection of methods which included teachers meeting weekly, providing detailed lesson plans for teachers, meeting with researcher researchers and unannounced lesson observations (Quinn, 2002).

1.11.10 Measures and dependent variables

All the PA studies included in the review collected quantitative data. One study (Vasileiadou, 2009) supplemented this data using observations to provide additional triangulation. Four of the studies used socio-metric methods, where for example children choose who they would most like to play with (Capodieci et al., 2016; Jacques

et al., 1998; Oortwijn et al., 2008; Putnam et al., 1996, Vasileiadou, 2009) with other studies using a combination of techniques (Piercy et al., 2002). Three studies provided details of the reliability and validity of the measures used (Capodieci et al., 2016; Jacques et al., 1998; Oortwijn et al., 2008).

Of the two bullying studies, one study used a bullying and victimization scale (Van Ryzin & Roseth, 2018) measuring externalising antisocial behaviour and negative behaviour during playground settings (Quinn, 2002). A summary of the variables and measures used can be found in Table 1.4.

Table 1. 5 Summary of variables and measures used in each study included in the review

Dependent Variables	Measures	
Capodieci, Rivetti, & Cornoldi (2016)	- ADHD symptoms - Social skills - Cooperative behaviour	Scala per i Disturbi di Attenzione/Iperattività per Insegnanti [ADHD scale for teachers]; Marzocchi & Cornoldi, 2000) Social Questionnaire, a supplementary section of the COM scale (Rivetti & Capodieci, in press).
Jacques, Wilton, & Townsend (1998)	- Social acceptance - Self-esteem	Sociometric measures (who would you like to invite to your birthday) Coopersmith Self-Esteem Inventory (Coopersmith, 1976)
Piercy, Wilton, & Townsend (2002)	- PA	PA scale adapted from (Moe, Nacoste and Insko, 1981)
Putnam et al (1996)	- PA - Student IQ - Student behaviour	The sociometric measure (Asher & Taylor, 1981) WISC-R Behaviour Rating Profiles (BRP, Brown & Hammill, 1983)
Oortwijn, Boekaerts, Vedder & Strijbos (2008).	- Popularity - Perceived non-cooperativeness	Teacher checklist on CL implementation Videotaped teacher-pupil interactions Mathematical ability Mathematical post-test Linguistic proficiency Coding of verbal interactions

Quinn (2002)	- Teacher nominations - Externalising anti-social behaviour - Observed behaviour - Peer Social Behaviour (PSB)	Teacher nominations Externalising antisocial behaviour Observed behaviour
Van Ryzin & Roseth (2018)	- Perceived stress - Emotional problems - Relatedness - Bullying and Victimisation - Engagement	Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983), 3 items from the Emotional Problems subscale of the Strengths and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998), 4 items from the Relatedness Scale subscales from the University of Illinois Bully Scale 4 items from the Behavioral Engagement subscale of the Engagement versus Disaffection with Learning Scale (Skinner & Belmont, 1993)
Vasileiadou (2009)	- PA	Individual semi-structured interviews Observations.

1.11.10.1 Relevant findings

All of the studies in the review, except Vasileiadou (2009), used inferential statistics such as Analysis of Variance (ANOVA), t-test and correlations to analyse the data. Only two studies reported effect sizes (Capodieci et al., 2016; Van Ryzin & Roseth, 2018)

Six studies looked at PA and related constructs such as popularity. Two studies looked at bullying or related concepts such as victimisation or negative playground interactions.

1.11.10.2 PA

All of the PA studies reported a positive impact on PA or a related construct following CL. Of the studies, five found significant increases in PA following the intervention. Jacques et al (1998) found significant social acceptance scores for the children with a 'mild intellectual disability' in the CL group compared to the control and 5 weeks later. However, Jacques et al (1998) did not report exact p values. Piercy et al (2002) found PA ($p < .004$) and popularity ratings ($p < .01$) of children with 'intellectual disabilities' from the children without disabilities increased significantly from pre to post test in the CL group but not in the comparison or control group (Piercy et al., 2002). Although *p* values were reported Piercy et al (2002) did not report effect sizes, therefore, the exact impact of the intervention is difficult to determine. In terms of the generalisability, two of the PA studies were focused on the outcomes for specific populations with children with 'mild intellectual disabilities' (Jacques et al., 1998b) and 'moderate to severe intellectual disabilities' (Piercy et al., 2002).

Putnam et al., (1996) found a significantly positive change in the desire to work with a classmate in CL conditions rather than in competitive conditions ($p < .001$). Perceptions of 'Special Education' students also became significantly more positive in CL condition than in the competitive conditions ($p < .001$). (Putnam et al., 1996). Oortwijn et al (2008) found popularity ratings increased ($p < .007$) along with a decrease in non-cooperativeness at the end of the CL intervention. However, there was no control for this so caution should be taken in attributing the results to the intervention due to potential confounding variables.

One study found a significant difference between the number of children with ADHD symptoms who were rejected by their peers, before and after the intervention (p =.03). However, the differences only approached significance for rejection as a 'desk mate' and 'teammate' (p =.07). Although not significant the data showed that a switch between rejected and accepted status occurred in the CL group and not in the control group (Capodieci et al., 2016). The final study reported 9 out of 10 children had an increase in the number of votes they received and 3 out 4 children who were considered to be rejected had an increased number of votes following CL, although this increase did not reach statistical significance (Vasileiadou, 2009). Vasileiadou (2009) did not include a control or comparison group and therefore results may be vulnerable to other factors affecting the reliability or validity of the findings. This contributed to a low WoE rating given to this study.

1.11.10.3 Bullying

In the bullying studies, only one study found significant differences. Van Ryzin & Roseth, (2018). found significantly lower levels of bullying and victimisation (p <.05) for marginalised students but not for other students. Quinn (2002) reported no significant difference in negative playground interactions or teachers' perceptions of externalising antisocial behaviour.

1.11.11 Strengths and limitations of the systematic review

The systematic literature review used a clear search strategy and an explicit inclusion criterion to identify studies. However, this may have meant that pertinent studies to the research area may have been overlooked if for example if they are unpublished, which may have had implications for the conclusions drawn. A WoE criterion allowed for the inclusion of different designs whilst placing less emphasis on designs which may not be the most effective for answering the research questions One important limitation to consider is that the review was completed by a single researcher and is therefore subjective and vulnerable to misinterpretation. The attribution of quality ratings and application of inclusion criteria is also a subjective process and multiple reviewers would be advisable (Petticrew and Roberts, 2006). In order to reduce subjectivity, an analysis of the weightings can be found in Table 1.2.

1.11.12 Summary and conclusions of the systematic review

The aim of the systematic literature review was to address the review question 'What is the impact of CL on pupils' PA, empathy and bullying behaviour? The systematic review highlighted that CL can have a positive impact on PA or a related construct. In relation to empathy the review highlighted that there is limited published research in relation to CL and empathy as no full-text articles were found during the search therefore conclusions regarding the impact cannot be made. In relation to bullying the review highlighted a potential positive effect of CL reducing bullying, however as with empathy the review highlighted that there is limited published research in this area with only two studies looking at bullying or related concepts. coo

The PA studies used sociometric measures to assess differences in PA before and after CL. All of the six studies all of the studies found positive effects of CL on PA. Two studies found that following CL mainstream pupils were more accepting of their peers who were considered to have 'intellectual disabilities' (Jacques et al., 1998; Piercy et al., 2002). Other studies showed peers were more likely to want to work with a classmate following CL (Capodieci et al., 2016; Putnam et al., 1996) with studies highlighting increases in popularity following a CL intervention (Oortwijn et al., 2008; Vasileiadou, 2009).

The number of studies found in the systematic review in relation to bullying were limited, which could be interpreted to reflect the need for further research in this area. The recent study (Van Ryzin & Roseth, 2018) included in the review did, however, highlight the potential positive effects of CL on bullying.

The majority of the CL interventions used Johnson and Johnson model with one study using Slavin's STAD model. A number of studies did offer extensive training (e.g. Putnam, 1993), however, the difficulty in implementing CL and the findings by Antil, Jenkins, Wayne, & Vadasy (1998) may provide some explanation as to why some researchers in the study conducted the intervention. In relation to the CL intervention, a further finding of note was that few studies included the pre-teaching of social skills within the intervention, something which is advocated within the literature as being an important component of CL (Goodwin, 1999).

The systematic review highlighted the lack of UK research therefore generalisability of the results may be limited. The review was further limited in the respect that some

studies did not include a control or comparison group, which makes it difficult to determine whether the changes would have occurred naturally without the intervention.

The review does, however, highlight the potential for CL to have a positive impact on PA. Although the results from the review in regard to bullying and particularly empathy are somewhat limited, the recent study on bullying highlights the potential for further research in this area. Given this, the conclusion of the review is that the evidence base for the impact of CL on PA and bullying and empathy is only starting to emerge within the literature. The review also highlighted a lack of research exploring the relationships between empathy, PA and bullying, despite this being a construct thought to underpin both PA and bullying.

1.11.13 Unpublished literature

Due to the lack of UK based research regarding the impact of CL on PA, bullying or empathy identified through the primary literature review a further search of the 'ProQuest Dissertations and Theses: UK & Ireland' database was conducted. A search was carried out on 20th July 2018 which followed the same search strategy as outlined previously without the criteria of papers being peer reviewed. The search resulted in 22 results. Titles and abstracts were screened in accordance with the inclusion criteria. As a result, 8 results were excluded from the review, 7 for not being related to any of the three variables, one because the methodology used was not experimental and one because the full text could not be obtained. An outline of the results and reasons for inclusion or exclusion can be found in Appendix D. Two studies remained which were related to PA (Craig, 2010; Smith, 2017).

1.11.13.1 PA

Craig (2010) evaluated how effective CL was in promoting PA levels of children within two year 4 classes in a mainstream primary school. The study used a pre-test post-test non-equivalent group quasi-experimental design. PA was measured by the Social Inclusion Survey (SIS) (Fredrickson & Furnham, 1998), which is a socio-metric measure, and the Strengths and Difficulties Questionnaire which looked at peer problems and prosocial behaviours. Craig (2010) analysed the data using inferential analysis and found children who received CL were, on average, significantly more accepted at post-test in both work and play contexts. In addition to this Craig found

significantly greater levels of ‘prosocial behaviour’ and reduced ‘peer problems’ compared to the control group. The CL intervention was based on Kagan approaches and included social skills training. Craig (2010) acknowledged limitations with the study related to intervention only being implemented by one class teacher in a single setting, which limits external validity.

Smith (2017) extended Craig's (2010) research increasing external validity by increasing the number of participants and contexts. Similarly, to Craig (2010) Smith (2017) used a pre-test post-test non-equivalent group quasi-experimental design to explore the impact of CL on pupils’ PA and peer support. Smith (2018) also used the SIS (Fredrickson & Furnham, 1998). Findings indicated that CL can enhance pupils’ perceptions of peer support, particularly in the work context and with pupils with lower levels of PA. However, as Smith (2017) noted the degree of impact was not consistent across individual classes, therefore, caution should be taken when interpreting the results.

1.12 The rationale for current research

The literature so far has highlighted the importance of PA for a range of academic, social and psychological outcomes. Learning structures within the classroom can implicitly communicate values (Johnson & Johnson, 2009; Johnson & Johnson, 2012). Teachers are situated just outside of the peer ecology and can help to shape the microsystems in which children are interacting (Rodkin & Gest, 2010) through being in a position to develop, guide and direct the classroom and act as an ‘invisible hand’ (Farmer, Mcauliffe Lines, & Hamm, 2011). For example, competitive learning structures implicitly communicate that success is contingent on beating and defeating others (Johnson & Johnson, 2009) which can foster negative interdependence, undermining desired social outcomes such as sharing, helping and empathy (Kagan & Kagan, 2015). In individualistic learning, students work alone to meet learning goals with no need for social interaction or interdependence which could promote the notion that success depends on a students’ efforts and is not positively influenced by others, with no need to develop positive relationships with peers.

By contrast, in CL groups work together to meet common learning goals, requiring positive interdependence and resulting in promotive interaction as individuals

encourage each other to reach the group's goals (Johnson & Johnson, 1992). This cooperation with peers is thought to result in increased empathy, altruism and perspective taking (R. Johnson & Johnson, 1992)and PA (Craig, 2010; Putnam et al., 1996; Smith, 2017).

Although links have been made between CL and bullying (see Van Ryzin & Roseth, 2018) there is limited research exploring this link experimentally with further research needed. Bullying has also been associated with low PA (de Bruyn et al., 2010) and low affective empathy (Jolliffe & Farrington, 2006), factors which are thought to be promoted by CL. As a result, in CL contexts it is expected students will develop a genuine sense of care and respect for their peers and as the classroom peer ecology changes and becomes less hierarchal power differences and opportunities for bullying would reduce (Hymel et al., 2015).

Therefore, this research will explore whether CL, through the facilitation of positive interdependence and subsequent changes in the classroom peer ecology can promote PA and empathy and improve pupils experience of bullying in school.

1.13 Research questions and hypotheses

The overarching aim of this project was to evaluate the effectiveness of a Cooperative Learning Intervention in enhancing PA, empathy and improving pupils experience of bullying in school. In order to achieve this a number of research questions had to be addressed.

Research Question 1: What is the impact of CL on pupils' PA?

Research Question 2: Does the impact of CL on PA differ depending on pupils' initial levels of PA?

Research Questions 3: What is the impact of CL on pupils' empathy

Research Question 4: What is the impact of CL on pupils' experience of bullying in school~~?~~

2 Methodology

This chapter outlines the methodology of the research. Different approaches to research in terms of ontology and epistemology are discussed before outlining the approach taken within the current research. The selection of measures used within the research will be outlined along with the rationale underpinning these. Ethical considerations and issues of reliability and validity will be discussed.

2.1 Approach to research

Researchers strive to search for what constitutes 'truth', although this holds different notions for different researchers. Research can be conceptualised within a number of different paradigms and many coexist within psychological and educational research (Cohen, Manion, & Morrison, 2011). A paradigm has been described as a loose collection of logically related assumptions which orientate thinking and research (Bogdan & Bilken, 2007) and consists of the following components: ontology, epistemology and methodology. Ontological assumptions are concerned with what constitutes reality (Scotland, 2012). Epistemology is concerned with the nature of knowledge (Scotland, 2012), how it is acquired and how it is communicated (Cohen, Manion & Morrison, 2011). The methodology refers to the strategy which is concerned with what, why, from where and how data is collected (Scotland, 2012). The guiding philosophy of research can, at times, be unclear but it remains of critical importance. As Scotland (2012) asserts, it is impossible to engage in any research without committing to ontological and epistemological positions as these influence the methodology and methods used.

Educational research has witnessed a proliferation of paradigms over time (Cohen, Manion and Morrison, 2011), as a result there are a number of theoretical paradigms discussed within the literature with different texts often using different terms. Four major paradigms, Positivism, Post-positivism, Interpretative/Constructivist and Pragmatic, which are felt to be most applicable to research within psychology and educational psychology, will be discussed.

2.1.1 Positivism

Positivism, sometimes referred to as 'science research', reflecting a deterministic philosophy (Mackenzie & Knipe, 2006), assumes the social world can be studied in a value-free way and causal explanations can be provided (Mertens & McLaughlin, 1995). The ontological position of positivism is realism, a view that a discoverable external reality exists independent of the research (Pring, 2000): this external reality can be measured, based on the assumption objective knowledge is possible. Positivism is therefore directed at explaining relationships, typically through correlation and experimentation, reducing complex interactions to their constituent parts (Scotland, 2012). These ontological and epistemological foundations have, however, been the focus of criticism, with critics arguing these positions are reductionist, with limited application to human behaviour (Mertens & McLaughlin, 2004) and to complex social worlds (Scotland, 2012). Some criticisms are addressed through the development of post-positivism.

2.1.2 Post-positivism

Post-positivism has similar ontological and epistemological positions to positivism. It, however, acknowledges the influence of theory, hypothesis, background knowledge and values on what is observed (Mertens & McLaughlin, 2004) and recognises that realities can be known only imperfectly and probabilistically (Robson & McCartan, 2016). Post-positivism is 'realist', but considers uncertainties and doubt (Robson & McCartan, 2016), and allows educational research to deviate from the assumptions of rigorous scientific methods, while still seeking to understand causal relationships (Scotland, 2012). Still employing predictions, with methods aligned with quantitative data, analysis involves descriptive and inferential statistics (Scotland, 2012).

2.1.3 Interpretive/constructivist

Interpretive or constructionist paradigms aim to understand human experience suggesting reality is socially constructed (Mertens & McLaughlin, 2004). The ontological position of interpretivism is relativism and its epistemology is subjective (Scotland, 2012). From this view, the social world must be understood from the standpoint of individuals within it. Interpretive methodologies therefore employ qualitative data collection in order to understand behaviour. The paradigms validity could be questioned based on its arguable limited transferability; with highly

contextualised qualitative data and interpretations based on subjective individual constructions (Scotland, 2012).

2.1.4 Pragmatic

In contrast, pragmatism places the research problem at the centre. As a result, data collection methods chosen are those that are likely to provide insights into the question with no preceding loyalty to any paradigm (Mackenzie & Knipe, 2006). Pragmatism therefore advocates the philosophical or methodological approaches which work best for the question (Robson & McCartan, 2016). Although often endorsed within the literature (see Burke Johnson & Onwuegbuzie, 2004), a pragmatic approach may produce less generalisable knowledge and results may be seen as vulnerable to the influence of the researchers' personal bias (Burke Johnson & Onwuegbuzie, 2004).

2.2 The paradigm of the proposed research

The proposed research attempts to isolate, control and manipulate variables in order to understand causal relationships. However, taking place in a naturalistic school environment potential extraneous factors are acknowledged. Therefore, a post-positivist standpoint was adopted.

2.3 Research design

The post-positivist standpoint proposed leads to fixed, typically quantitative experimental and quasi-experimental designs; the latter without the randomisation of participants.

2.3.1 Research design selected

This research took place in a school, and random allocation was not appropriate or practical. The proposed research therefore employed a quasi-experimental pre-test post-test non-equivalent groups design. This was carried out across 2 schools, and 2 age groups, with an intervention and Comparison group. Groups were mixed, varying by ability and sex. Both groups completed three pre-test measures. The experimental group were introduced to CL structures whilst the Comparison group did not receive special treatment. Post-test measures were then taken from both groups following the intervention.

2.3.1.1 Independent variable

The independent variable for this study was the method of teaching, with two levels:

1) Cooperative Learning structures employed (intervention)

2) Same lesson content but with no cooperative learning structures employed (Comparison)

2.3.1.2 Dependent variables

In this study the dependent variables (measures discussed in section 2.7) were:

1) Peer Acceptance

2) Empathy

3) Pupils experience of bullying

2.3.2 Quasi-experimental designs

Quasi-experiments aim to infer causality between interventions and outcomes (Harris et al., 2006) although assignment to conditions is by self-selection (Shadish, Cook, & Campbell, 2002b). For the current research, the randomisation was impractical because the groups were dictated by class groupings. Therefore, although the control group was similar to the intervention group it could not be matched nor randomised. Despite this, quasi-experimental designs which employ control and experimental conditions with pre and post-testing in both conditions are considered in high regard (Harris et al., 2006).

2.3.3 Limitations of quasi-experimental designs

Quasi-experimental designs may hold the potential for non-equivalence between groups, with implications for the reliability and validity of data. In order to identify causal inferences, therefore, possible alternative explanations which could falsify the findings need to be explored such as threats to internal validity. (Shadish et al., 2002b).

2.3.4 Alternative designs considered

Alternative designs were considered including randomised control rrials (RCT's) single case experimental designs and non-experimental but were discounted for reasons of practicality, desirability or efficiency.

2.4 Stakeholder involvement

During the planning and implementation, key stakeholders were considered. These included:

- the University of Nottingham, as the research was conducted as part of a requirement for the researcher's doctoral training programme (DAppEdPsy).
- the Local Authority (LA) where the researcher was on a professional placement where the research took place and
- the staff and pupils at the schools where the research was carried out.

2.4.1 Selection of schools

Through discussions with EPs within the LA a number of potential schools were identified that met the research requirements. Contact was made through the researcher to the headteachers asking for expressions of interest in the research. Two schools responded. The researcher spoke with school managers to share information about the research and discuss the implications of being involved. An information sheet was provided, a copy of which can be found in Appendix E and an inclusion criterion was used to aid sampling which was as follows:

- Year 4 or 5 classes with a parallel class in the same year group;
- Mixed class groupings – grouping not based on ability or chronological age;
- Class predominately taught by class teacher (may be taught in sets for some lessons, e.g. maths);
- Both classes planning to deliver the same curriculum content during the chosen school term;
- Teachers available to attend training at one location;
- Teachers were willing and able to deliver the intervention in the Autumn Term; and
- Teachers having very little experience of CL structures.

Once suitable classes were selected consent was sought from the class teachers, which can be found in Appendix F. Letters were sent to parents via the usual school procedures. The letter explained the research and requested consent for their child to complete the measures. A copy can be found in Appendix G.

2.4.2 School characteristics

Two mainstream primary schools were recruited to take part in the research. Information from Ofsted described both schools as larger than average. In the schools' recent inspections, School A was judged as 'requires improvement' and School B was judged as 'good'. Specific school information is provided in table 3.1. The number of pupils with SEN in the schools was average and slightly below average respectively. In school A the number of pupils entitled to pupil premium was above average but was below average in School B. In both schools most pupils are described as being White British with English as their first language.

Table 2. 1 Key characteristics of the two participating schools.

School	**School A**	**School B**
Size of School	Larger than Average	Larger than Average
Category and Age Range	Primary 4 to 11	Primary 3 to 11
The proportion of pupils on the SEN register or with an Educational Health and Care plan (EHCP)	Average	Below Average
The proportion of pupils entitled to the pupil premium	Above Average	Below Average
The proportion of pupils from ethnic minority backgrounds	Below Average	Below Average
The proportion of pupils who speak English as an Additional Language	Below Average	Below Average
Ofsted grading and date of inspection	Requires improvement (January 2018)	Good (July 2012)

2.5 Participants

2.5.1 Sampling strategy

Participants were a convenience sample of pupils in Year 4 or 5 classrooms within the schools. The headteacher of a school was responsible for deciding whether Year 4 or 5 classes would take part in the intervention. This age range was chosen for three reasons. First, theories of development suggest between the ages of 7 and 11 years is the period where children should learn to cooperate with their peers and adults and begin to adjust to the needs of others in pursuit of mutual interest (Eccles, 1999). Second, research suggests CL yields the best results with students from primary level aged 6-12 years (Kyndt et al., 2013). It is also argued that junior school is an appropriate time to intervene as by this age problems of sociometric rejection and bullying become more identifiable (Cowie, Smith, Boulton, & Laver, 1994). It is therefore argued that prior to the transition to high school is a crucial time for bullying pre-intervention efforts (Hymel et al., 2015). Finally, pragmatically it was felt Year 3 pupils would still be settling into their new Key Stage and the examination demands on teachers and Year 6 pupils could interfere with the implementation of the research.

In School A the Year 5 class and in School B both Year 4 and 5 classes took part in the intervention. Once selected, class allocation to intervention or control was determined through discussion with the headteacher, based on teacher capacity and external pressures and willingness to take part. For example, a control group teacher in School B had not started at the school at the time of initial discussions and it was felt they would need time to become familiar with the school processes as a new member of staff. Teachers remained with their usual class and existing class grouping remained intact.

2.5.2 Class teacher characteristics

There were three female intervention class teachers. All the intervention teachers had 6 years of teaching experience. In the waitlist control condition, there were two female and one male teacher. Two teachers had taught for 15-16 years and one teacher 18 months. A summary of the teacher characteristics for each class is presented in Table 2.2.

Table 2. 2 Class teacher characteristics (intervention and control)

School	Condition	Teacher Gender	Number of years teaching
School A	Year 5 Intervention	Female	6
	Year 5 Control	Female	1.5
School B	Year 4 Intervention	Female	6
	Year 4 Control	Male	15
	Year 5 Intervention	Female	6
	Year 5 Control	Female	16

2.5.3 Class characteristics

Pupils were aged between 8 and 10 years old. Parental consent for their children to complete the measures was gained through an opt-in process, in accordance with ethical guidance. Information on each class (year group, number of pupils, gender and percentage of class with consent) can be found in Table 3.3. and Table 3.4.

Table 2. 3 A table to show the class characteristics for each school and condition providing details as to the total number of pupils in each class, gender and number of pupils with and without consent

School	Condition	Year Group	Total Number in Class	Gender M/F	Number of pupils with consent	% of pupils with consent	No consent
School A	Intervention	5	24	M 13 11 F	23	95.8%	1
	Control	5	24	M 12 F 12	20	83.3%	4

School B	Intervention	4	29	M 17 F 12	25	86.2%	4
	Control	4	28	M 18 F 10	23	82.1%	5
	Intervention	5	27	M 16 F 11	24	88.9%	3
	Control	5	26	M 12 F 14	23	88.5%	3

Table 2. 4 The total number of participants in intervention and control conditions, including information on consent levels.

	Intervention	**Control**	**Total N**
Number with consent	72 (37m/32f)	66 (28m/32f)	138
Number without consent	8 (7m/1f)	12 (10m/2f	20
Total N	80	78	158

In total there were 80 pupils who received CL and 78 pupils who were part of the Comparison group. All children (n=158) were rated as part of the Peer Acceptance

measure. From this 87% of pupils (n= 138) had consented to complete the measures. Out of the 138 pupils who consented five pupils were absent on the day of pre-data collection and 6 pupils chose not to complete the questionnaires (Total n= 127; 63m/64f). Out of the 127 pupils who completed pre-questionnaires 10 pupils were absent on the day of post data collection (6f/4m) therefore the number of participants who completed both pre and post questionnaires was 117 (59m/58f).

2.6 Procedure

2.6.1 Pilot

The researcher attempted to undertake a pilot study in a different school in the summer term prior to intervention beginning in the Autumn term to test the suitability of the measures and training materials. However, due to difficulties in staff availability it was not possible to polit the training materials and difficulties obtaining parental consent meant the measures were only piloted with a small group of children. This led to subsequent changes in one the measures as discussed in section 2.7.2.1.

2.6.2 Training

Intervention class teachers received a 2-hour twilight school-based training session in CL, which was led by the researcher. The Kagan model was selected, since literature indicates that more direct approaches, such as Kagan's model, require less training and are more easily implemented (Johnson & Johnson, 2002). The training was developed by the researcher with guidance taken from previous doctoral research on CL (Smith, 2017). The session included:

- An introduction to CL which focussed on what CL is, its history and primary psychological theories which underpin CL and relevant research in terms of social, emotional and academic outcomes.
- The teaching of social skills using T-charts
- The structuring of CL groups (e.g. a maximum of four pupils and teacher selected, heterogeneous groupings).
- Training on five main Kagan structures with opportunities for teachers to practice these structures within the training session through activities.
    - Timed Pair Share

    - Numbered Heads Together
    - Quiz Trade
    - Inside Outside Circle
    - Robin Structures (including round robin and rally robin)
- Frequently Asked Questions about Cooperative Learning

At the end of the training. research questions and expectations were clarified. Teachers were provided with copies of the training material in the form of a resource booklet, which provided key information in relation to CL and the structures to be implemented in the classroom. Extracts from the training materials can be found in Appendix H. At the end of the training there was time dedicated to clarifying the research and the expectations.

2.6.3 Coaching sessions

In addition to the training, four coaching sessions took place within the intervention period. The rationale for these originates from the literature suggesting potential challenges in the implementation of CL, with it often being abandoned or reduced following formal training (Sharan, 2010). As Sharan (2010) states the change in the role of the teacher and change in the interaction between students and teacher involved in CL 'requires commitment, repeated practice and a network of support, encouragement and feedback' (p.306). In order to expect teachers to adopt a teaching practice attention needs to be paid to learning and behavioural change processes that lead to different teacher behaviours and improved student outcomes (Truscott et al., 2012). Elements from the Exceptional Professional Learning (EPL) (Truscott et al, 2012) were used to develop a structure for the coaching sessions.

The EPL provides a coordinated framework to elicit and sustain change in cognition and behaviour among consultees. Although the EPL is advocated for use over a longer period of time the researcher considered some of the theoretical groundings of the framework was useful to consider when planning coaching sessions. Within the training, activities were included to allow teachers to experience CL as this is often seen as a key feature of professional learning (Truscott et al., 2012). Pragmatically the initial training needed to be delivered in one session. However, the researcher recognised that this massed practice approach was likely to be less effective as research has shown the distributed practice is more effective than massed practice (Truscott et

al., 2012). The coaching sessions, therefore, allowed more opportunities to practice and for teachers to reflect and problem solve.

2.6.4 Structure of the coaching sessions

The EPL framework was used as a guiding structure for the sessions. However, it was recognised that teachers may have areas which they would like to discuss. The first part of the session aimed to draw upon the EPL component of 'Relatedness' to develop feelings of connectedness with the other members of the group (Ryan & Deci, 2000). This was achieved through opportunities for discussion and group problem-solving. The next part of the session aimed to draw upon the 'Authenticity' component and genuine and authentic applications of practical content in the setting. Teachers were asked to share how confident they felt and discuss examples of how they have implemented the structures in the lessons. Some collaborative discussions were also held during and following fidelity checks about the lesson and structures observed. Teachers were asked to reflect upon contextual factors which may be interfering with the implementation of the structures and were given an opportunity to practice structures or collaboratively plan lessons if needed.

The process of reflection is considered important as it supports teachers in dealing with the complexity of CL, can inform practice and facilitate its transformation (Sharan, 2010). The next part of the session aimed to focus on the 'Autonomy' component with the aim to build autonomy through reflecting on successful practice and actively engaging teachers in decision making, problem-solving and goal setting. The next component, 'responsiveness', was addressed by having an evolving agenda and having time to recap over key principles or structures if needed. The last component, "equity", was achieved by making clear the knowledge that the teachers have and its importance in being able to implement CL. Positive feedback and an acknowledgement of strengths were key elements of the coaching session. A template used to guide the coaching sessions can be seen in Appendix I.

2.6.5 Cooperative learning intervention

The intervention ran for a period of 11 weeks, which included an initial week of social skills training. The intervention period was split with a one-week half-term holiday. During the intervention, teachers were asked to implement CL approaches in at least four lessons per week, as research has suggested in order to bring a teaching model of

medium complexity under control requires about 20-25 trials in the classroom over a period of 8-10 weeks (Joyce & Showers, 2002). The CL sessions were delivered by the usual class teacher on a whole class basis and class groups remained intact. During the training, teachers were reminded not to use CL when teaching mixed class groups. In School B, children were in ability groups for Maths and English and therefore some of the class were being taught by the control teacher. The intervention teachers consequently only implemented the CL structures outside of these subjects. The implications of this will be discussed in the limitations of the research in section 4.3. In school A, children were in the same class for all of the lessons.

2.6.6 Comparison

During the intervention period, the Comparison classes received teaching as usual and remained in their usual groups with their usual class teachers. In School B children were taught by the intervention teacher for some lessons, however, the teacher did not use CL during these lessons. Pupils in the Comparison classes, therefore, received no intervention during this period. Following the completion of the research, the teachers in the Comparison classes were provided with the opportunity to receive the same training as the intervention teachers in order that all pupils had access to CL.

2.6.7 Intervention fidelity

Intervention fidelity refers to the extent to which core components of interventions are delivered as intended (Gearing et al., 2010). Intervention fidelity is crucial to judge the quality of a study (Murphy & Gutman, 2012) as it cannot be assumed that implementers will administer interventions flawlessly (Robson & McCartan, 2016). There is a broad agreement that intervention fidelity is a multi-dimensional construct (Durlak, 2010) composed of at least five intervention fidelity dimensions. The most commonly measured dimensions of intervention fidelity are dosage and adherence (Mendive et al., 2016). Dosage refers to how much of the programme was actually delivered and adherence refers to the level of correspondence between the intended intervention and the versions implemented (Dane & Schneider, 1998).

2.6.7.1 Adherence

Intervention adherence has often been measured using checklists to capture the presence or absence of elements targeted by a program (Hamre et al., 2010). The checklist then produces a score which represents the proportion of elements delivered. Literature encourages direct observation rather than teacher reports as it is thought to be more accurate and less biased (Hamre et al., 2010; Schulte, Easton, & Parker, 2009). Intervention adherence was particularly important given that the intervention was implemented across two schools, which poses a threat to validity. Four fidelity checks were scheduled into the research timetable.

Intervention adherence was assessed using a checklist which included the key components of CL. A copy of the checklist used can be found in Appendix J. Adherence checklists used in previous doctoral research into CL were used as a basis (see (Craig, 2010; Smith, 2017). Using the checklist, a judgement was made as to whether the component was clearly present, whether there was only some evidence or whether it was absent. Each judgement was scored. If the component was present a score of 2 was given and if there was some evidence of the component then a score of 1 was given. If there was no evidence a score of 0 was given. Similar to the approach used by Smith (2017) a double weighting was given to the core principles of CL, that is, positive interdependence, equal participation, individual accountability and simultaneous interaction (Kagan & Kagan, 2015), along with the presence and reflection on small group skills. It should be noted that inter-rater reliability checks

were not carried out. The implications of this will be discussed within the review of the research in section 4.2.3.

2.6.7.2 Dosage

Dosage is typically measured by counting the time devoted to the intervention. Intervention dosage was recorded by each intervention class teacher completing a daily record log. This was used to record the frequency, duration and type of CL structures used within the lesson.

2.6.8 Diffusion of treatment

Two fidelity checks were planned in the control groups at the beginning and end of the research to identify whether any CL elements were being used and to establish whether there was any likely diffusion of treatment. Due to staff absences not all of the control group fidelity checks were carried out. Compared to the intervention group the control groups scored low on the fidelity checks with nearly all CL elements absent. In school B the percentage increased from check 1 to check 2 which could indicate some diffusion of treatment through conditions. Scores for fidelity checks carried out in the control group can be found in Table

Table 2. 5 Intervention adherence: raw scores and percentages in waitlist control condition.

Comparison class	**Intervention Adherence**				**Mean raw score**	
	Check 1		**Check 2**		**Mean %**	
	Raw	**%**	**Raw**	**%**		
School A Year 5	4	9%	5	11%	5	10%
School B Year 4	6	13%	10	22%	8	18%
School B Year 5	5	11%	n/a	n/a	5	11%

2.7 Measures

To be able to answer the research questions, three dependent variable measures were used:

- Social Inclusion Survey (SIS) was used to measure peer acceptance (Fredrickson and Furnham, 1998).
- Empathy Measure (Adapted from the Emotional Literacy Scales) (Faupel, 2003)
- My Life in School Checklist was used to measure a pupils' experience of bullying (Arora & Thompson, 1987).

All measures used were pupil self-report, as the primary focus of the research was to establish whether there was a perceived change in the pupils themselves. Potential difficulties with self-report are acknowledged, which could include social desirability bias (Van De Mortel, 2008).

2.7.1 Measuring peer acceptance: The Social Inclusion Survey (SIS)

The Social Inclusion Survey (SIS) (Fredrickson and Furnham,1998) was used as a measure of peer acceptance within the classroom. The SIS is a sociometric scale designed to be administered on a whole class basis to children above 7 years old (Frederickson & Furnham, 1998). It is made up of two questionnaires, Like to Work With (LTWW) and Like to Play With (LTPW).

There is space for the names of each pupil in the class, next to which there are four circles, one containing a question mark and the others containing a schematic face which displays a smiling, a sad and a neutral expression (Appendix K). On the LTWW questionnaires pupils are asked to tick the face which shows how much they like to 'work with' each person in their class: the smiling face to indicate which classmates with whom they like to work; the straight mouthed face to indicate classmates whom they do not mind whether they work with or not; and the sad face to indicate classmates whom they prefer not to work with. Pupils are asked to use the question mark category to indicate any classmates they do not know well enough to decide. The LTPW questionnaire follows an identical format and instructions with the exception that children are asked who they like to 'play with' rather than 'work with'.

2.7.1.1 The rationale for using the Social Inclusion Survey

The SIS (Fredrickson & Furnham, 1998) was selected as a measure of peer acceptance because it offers an unobtrusive means by which pupils' social inclusion can be monitored (Fredrickson &Furnham, 1998). The forced choice techniques mean the information was gained about each pupil in the class. Additionally, the SIS has demonstrated high test-retest reliability scores (0.70-0.78) for pupils aged 9-12 years over a 5-week period (Fredrickson and Furnham, 1998b). The SIS has also been used in previous doctoral research exploring peer acceptance (Craig, 2010; Smith, 2017).

2.7.1.2 Scoring

To score the SIS, the questionnaires for each class were organised according to the sex of participants as Fredrickson and Furnham (1998) states it is appropriate to regard 'same-sex and opposite-sex peer groups as interrelated but discrete social subsystems within the classroom' (p.930). As suggested by Fredrickson and Furnham's (1998) instructions the numbers of smiling, sad and neutral faces are tallied and entered separately for the same sex, the opposite sex and whole/class group. This procedure was followed for both 'work and 'play' questionnaires.

Fredrickson and Furnham's (1998) guidelines suggest categorising pupils as 'rejected, average or popular', however previous doctoral research has rejected this scoring approach, as it only produces ordinal level data, in favour of a weighted approach (Craig, 2010; Smith, 2017). The use of ordinal data and the benefits of rescaling this into interval data have also been discussed in the literature (see Harwell & Gatti, 2001). Therefore, in line with previous research, to aid continuity and generalisability, a weighted approach was used. A 'weighted score' was produced for each participant for same-sex and opposite-sex ratings by calculating the total number of smiling faces ticked for the participant and subtracting the number of sad faces. Neutral faces and any spoiled responses were counted as zero. To allow for a comparison to be carried out between classes of different sizes and numbers of male and female participants, this number was then divided by the total number of respondents. The weighted score approach enables interval level data to be produced which allows for more sophisticated analysis.

Each class member received four peer acceptance scores ranging from -100 (low peer acceptance) to +100 (high peer acceptance).

- work same gender (WSG)
- work different gender (WDG)
- play same gender (PSG)
- play different gender (PDG)

For example, if a male pupil received 6 happy faces, 3 neutral faces and 2 sad faces from 11 male peers the LTWW questionnaire. This peer acceptance score would be calculated as 6 happy faces minus 2 sad faces, giving a score of 4. This score would then be divided by 11 (number of respondents) and multiplied by 100. This would give a final work same gender (WSG) peer acceptance score of 36.36.

2.7.2 Empathy measure

The instrument used to measure empathy was adapted from the Emotional Literacy Scales pupil checklist (Faupel, 2003). The Emotional Literacy Scales pupil checklist is a widely used self-report measure for students aged 7-11 which looks at aspects of self-awareness, self-regulation, motivation, empathy and social skills which are combined to give an overall emotional literacy score. Only the 5 items from the empathy factor were used along with 5 items which acted as distraction items. A copy of the measure can be found in Appendix M..

2.7.2.1 The rationale for using the empathy measure

The researcher attempted to locate an appropriate empathy measure within the literature which had established good reliability and validity. This revealed a limited number of measures available to assess empathy within children. These were often difficult to obtain.

Following a detailed exploration of the literature, the Kids Empathy Developmental Scale (KEDS) (Reid et al, 2013) was selected as a potential measure. The KEDS is a measure of empathy asking pupils to look at picture-based scenarios and infer what emotions that person or persons in the image might be feeling, and attributing reasons for these feelings. The KEDS had demonstrated good internal consistency measured by the Cronbach's alpha on the affect scale (.63), cognitive scale (.82) and behavioural scale (.80). The measure was piloted with a small group of pupils and this highlighted some potential questions around its appropriateness. The KEDS was made up of 13 picture-based scenarios and for each picture there was a minimum of four questions

with some items having additional questions items. In total there were 64 questions. Following this piloting it was decided that the measure was not appropriate as 1) it took a significant amount of time to administer, 2) it is recommended to be administered individually and therefore would not be suitable for whole class administration and 3) the scenarios used did not seem to reflect current social contexts.

After consideration, an alternative measure was explored and after comparison of available measures and further difficulties obtaining a suitable empathy measure, it was felt the Emotional Literacy Scales pupil checklist could be utilised (Faupel, 2003). The Emotional Literacy Scales pupils' checklist has been standardised and has reported an overall reliability of 0.76. This full checklist has 25 items and due to pupils having to complete a further two measures it was decided that only the empathy factors would be used. However, the reliability for the empathy dimensions is lower so the following steps were taken:

- Reliability coefficients were recalculated (see section 3.15)
- Only descriptive data is presented
- Triangulation between this and other measures
- A discussion on the lack of well-attuned measures of empathy

The ethical implications of using an alternative measure were considered and an ethics amendment was drafted (see Appendix N). However, consultation with the ethics procedures was undertaken and this states if the change poses no ethical risk, then further ethical approval is not required.

2.7.3 Measuring the experience of bullying in school

The My Life in School Checklist, junior version (Arora & Thompson, 1987) was used to measure pupils experience of bullying. This checklist is a 40-item questionnaire which includes a mix of statements describing bullying, friendly and aggressive behaviour. Pupils are asked to indicate whether they had experienced the situation 'not at all', 'only once' or 'more than once' during the last week of school. Six key items from the questionnaire are used to calculate a Bullying Index and General Aggression Index. These items are 'tried to kick me', 'said they'd beat me up', 'tried to make me give them money', 'tried to hurt me', 'tried to break something of mine' and 'tried to hit me'. A pilot of the measure highlighted that the questionnaire was quite long

therefore the number of items on the questionnaire was reduced from 40 to 15 items. This is in line with the measure guidelines which states that items can be added or removed, without affecting the reliability, as long as the key items for calculating the Bulling Index and General Aggression Index are not altered.. A copy of the measure can be found in Appendix L.

2.7.3.1 The rationale for using My life in School Checklist

The My Life in School Checklist can be used to gauge the extent of bullying behaviour, aggressive behaviour and friendly behaviour at a particular point in time. The questionnaire can be adapted, and items can be added or removed as long as key items for calculating the Bullying Index and the General Aggression Index remain unaltered. A split-half reliability test carried out on 51 pupils from year 8 to year 9 indicated good internal reliability. The checklist has been used in previous doctoral research to measure reported levels of bullying (see Herrick, 2012).

2.7.3.2 Scoring of the My Life in School Checklist

Scoring was completed in accordance with the guidelines (Arora & Thompson, 1987). The Bullying index was calculated using the following steps:

Step 1: For each of the six key items on all questionnaires the number of times a tick was placed in the 'more than once' category was counted.

Step 2: For each of the six items the scores were divided by the number checklists completed and multiplied by 100. This gives a percentage of pupils who responded with 'more than once' for each item.

Step3**:** The six percentages were added together and divided by six.

The General Aggression Index is calculated following steps 1 and 2 above then:

Step 3: For each of the six items on all questionnaires, the number of times a tick was placed in the 'once' category was calculated.

Step 4: For each of the six items scores were divided by the number of checklists complete and multiplied by 100 to give a percentage of pupils who responded with 'once' for each item.

Step 5: The 12 percentages (six from the Bullying Index and six from the General Aggression Index) were added together and divided by 12.

2.8 Data collection procedures

Pupils completed each measure at pre and post-test. Pre-measures were taken in the week prior to the intervention and post measures taken in the week immediately following the end of the intervention.

2.8.1 Procedure for administering measures

The pupil information sheet was presented electronically on the interactive whiteboard and read out by the researcher (see Appendix O) Paper copies were available if the children wanted to take one. If parental consent had been received, pupils completed a 'pupil consent form' indicating whether they were or were not willing to complete the measures, a copy of which can be found in Appendix P.

Each measure was administered on a whole class basis by the researcher. The SIS was completed first, then the My Life in School Checklist followed by the empathy measure. Teaching staff remained in the classroom and were available to support any children who encountered difficulties in completing the questionnaires. The teachers provided the children who did not have parental consent to complete the measures with an alternative activity.

A researcher-designed script was used to ensure the standardised delivery of the measure across the classes and schools to reduce researcher bias. On the SIS (Fredrickson & Furnham, 1998) and My Life in School Checklist (Arora & Thompson, 1987) children worked at their own speed but were told to ask an adult if there was something they were unable to read or were unsure of. For the empathy measure, the researcher read out each item and asked children to work at the same speed to ensure children had a good understanding of the items.

In order to make attempts to reduce bias in self-reports, pupils were reminded to be honest and that there were no right or wrong answers. Pupils were told that their answers will not be shared, and their name will not be on the questionnaires. Pupils were asked not to share their responses with other pupils in the class explaining it

might upset children if they knew another pupil had put a sad face next to their name. Some children chose to use a book to hide their answers.

2.8.2 Anecdotal Data

Anecdotal data was gathered and recorded throughout the research process from observations from teacher and pupil comments during fidelity checks, teacher comments on record sheets, and teacher comments and feedback during coaching sessions. A summary of this data, which includes exerts from teachers, is presented in section 3.18.53.18.5 and is used to compliment the discussion of the findings and future research. The anecdotal data gathered is specific to the teachers in the research and therefore its limitations in terms of reliability and validity are acknowledged.

2.9 Ethical considerations

Key ethical guidelines from the British Psychological Society (BPS) (BPS, 2010; BPS, 2002) and the Health and Care Professions Council (HCPC, 2016) were considered and ethical approval was granted from the University of Nottingham Ethics Committee (see appendix Q). The ethical considerations felt pertinent to the proposed research are outlined along with steps which would be taken to minimise the concern.

Informed consent - CL is viewed as an activity which would fall within the usual curriculum as the BPS (2010) Code of Human Research Ethics states that if research procedures are judged to fall within the range of usual curriculum or other instructional activities consent from the participants and granting of approval from a senior member of staff should be sufficient. Opt-in consent was gained from parents to allow their child to complete the measures. Children for whom consent was not obtained did not complete the measures. Pupils also provided consent before completing the measures. Pupils were allowed to opt out of completing the measures even if parental consent was sought.

Right to withdraw - School staff, parents and pupils were made aware that participation is voluntary and that they are able to withdraw from the research at any point, without having to provide a reason. It was made clear to pupils that consent could be withdrawn retrospectively and that any request for data to be destroyed was complied with.

Confidentiality - Data was reported anonymously to reduce the likelihood of individual pupils, classes or schools being identified and individual responses on the measures were not shared with staff.

Protection from harm - Completion of the measures was deemed unlikely to change participants mood or be stressful as the literature suggests a focus on peer acceptance and relationships does not impact negatively on children (Iverson, Barton & Iverson, 1997). However, it is recognised that the measure may have the potential to increase negative behaviour towards disliked peers. Steps were taken to highlight the importance of not discussing answers with peers.

Deception - The research purpose was explained to schools, parents and pupils from the outset.

Debriefing - As part of the debriefing procedures staff were asked to remind classes about the school anti-bullying policy and to speak to a member of staff if they felt they or other pupils were experiencing bullying. A copy of information provided to parents and pupils can be found in Appendix R and Appendix S. Following completion of the research schools, parents and pupils will be provided with a letter outlining the key findings of the research.

2.10 Reliability and validity

Validity refers to the accuracy of the result, which is the extent to which what is measured is what is claimed to be measured (Robson & McCartan, 2016). Validity in quantitative designs can be explored as internal and external validity.

2.10.1 Internal validity

Internal validity refers to the extent to which findings are related to the phenomena being investigated (Winter, 2000) and can include the following:

History – Changes over time due to exposure to external events**.** Discussions with stakeholders prior to or following the intervention did not highlight any school-wide changes or additional interventions taking place which could account for the changes in the Dependent variable.

Testing – Changes occurring due to those other than experimental treatments. The completion of the social inclusion survey could have led participants to focus on peer relationships following the pre-tests which could cause changes in post-testing. This threat was addressed using a control group.

Instrumentation – Changes in administration methods during pre and post-testing. A pre-developed script and visual presentation were used to ensure identical administration procedures were used for both the experimental and control group. All testing for all groups was completed within morning lessons.

Selection – Differences between the groups prior to involvement. The lack of randomisation means that groups may have been different prior to the study. As a result, inferences about causality are less certain. This threat will remain a limitation and will be discussed further in section 4.2.1.1.

Diffusion of treatment – this refers to when one group inadvertently learns information or receives aspects of the intervention only intended for the experimental group. There was a risk of the skills of the intervention group transferring to the control condition, particularly in School B as some of the classes were mixed for some lessons. To eliminate diffusion of treatment teachers were asked not to discuss the structures with the control group teacher and complete their planning separately.

2.10.2 External validity

External validity refers to the degree to which results can be generalised to a wider population, settings and situations (Cohen et al., 2011).

External validity was raised by using pre-existing classes, their usual teachers and the usual curriculum. The research was conducted within mainstream Local Authority schools, with realistic levels of support and training. Although the research was conducted within two schools across three-year groups, which reduces the chances that findings are associated with teacher characteristics, the population and context are restricted to a two-form entry school with year 4 and 5 participants.

2.10.3 Issues of reliability

Reliability is related to dependability consistency and replicability (Cohen et al., 2011).

The issues of participant bias were considered. Participant bias refers to participants responding in ways in which they think are expected or desired. In order to reduce the risk that social desirability effects may affect participants ratings, attempts were made to emphasise the confidential nature of the questionnaires and it was made clear answers will not be shared with staff or students. During the administration of the questionnaires, students were told not to share or look at answers so pupils could answer confidentially.

The measures were administrated by the researcher and a script was followed which is provided in Appendix N. to ensure that the measures are administrated consistently across conditions and across pre and post-testing. Observer bias is reduced by using self-report measures which meant no researcher interpretation was required.

3 Results

This chapter outlines the findings of the study and the statistical analysis used in judging the significance of the results. For each of the measures the mean scores are presented followed by statistical analysis and key findings.

3.1 Parametric and non-parametric tests

Before a data set can be inferentially analysed decisions need to be made about whether parametric or non-parametric tests should be used. Parametric tests make assumptions about the population from which the data has been from (Dancey & Reidy, 2007). These are described below:

1. The population from which the samples are drawn are normally distributed.
2. The variances of the population being compared are approximately equal.
3. There are no extreme scores within the data set as parametric tests can involve the calculation of the mean as a measure of central tendency and therefore extreme scores can distort this.

(Dancey & Reidy, 2007)

Parametric tests are preferred in psychological research as they have greater statistical power when compared to their non-parametric equivalents. However, if parametric assumptions are not met then typically non-parametric tests would be used. To ascertain whether the assumptions were satisfied approaches the following approaches were used:

- Consideration of skewness and kurtosis
- Tests of normality using Kolmogorov-Smirnoff and Shapiro-Wilk (Ghasemi & Zahediasl, 2012). The Shapiro-Wilk test was used for small sample sizes of less than 50 (Razali & Wah, 2011).
- Levee's test of variance
- Visual analysis of histograms and box plots to identify any extreme outliers.

3.1.1 Outliers

Outliers can increase variance error and reduce the power of statistical tests and decrease normality (Osborne, 2014). Although there is debate as to what possible

treatments of identified outliers including transforming the data (Wright, 2003) outliers were kept in their original form as it was felt they would be interesting in their own right. Non-parametric tests were used when the data sets included multiple outliers or included one particular extreme outlier.

3.2 Test Selection

3.2.1 Parametric test selection

Pre-test and post-test control designs are commonly used in educational research (Dugard & Todman, 2006). However, despite this popularity, there is long-standing controversy regarding the analyses of data, with conflicting opinions as to whether an independent t-test on 'gain' scores (post-test minus pre-test) or an analysis of covariance (ANCOVA) (Knapp & Schafer, 2009) is more appropriate.

Debate centres around the claim that both methods can yield disparate results, which has come to be known as the 'Lords Paradox' (Knapp & Schafer, 2009). Gain score analysis is criticised for being deceiving in its simplicity and it has been argued to be less reliable (Dimitrov & Rumrill, 2003; Knapp & Schafer, 2009). Despite this criticism an ANOVA or independent t-test is argued to be less biased in non-randomised studies, when baselines differences between groups exist. (Allison, 2006; Oakes & Feldman, 2001; Van Breukelen, 2006). Therefore, when parametric assumptions were met, t-tests or one-way ANOVA's on gain scores were used.

3.2.2 Non-parametric Selection

In contrast, non-parametric tests do not have such stringent requirements and do not make assumptions about the normality of the data (Harwell, 1988). Most psychological and educational data are quite flat or skewed and non-parametric tests are typically therefore considered more appropriate on a statistical basis (Harwell, 1988). When parametric assumptions were not met, in this study non-parametric statistical analyses were conducted using the Mann-Whitney U, Wilcoxon Signed Rank and Kruskal-Wallis test.

3.3 Effect size

Effect sizes quantify the effectiveness of an intervention relative to a comparison group (Coe, 2002) and are typically interpreted based on rules of thumb, for example, Cohen (1998) considers an effect size of approximately 0.20 to be 'small', approximately 0.50 as 'medium' and approximately 0.80 is considered to be a large effect size. The use of such distinctions is commonplace amongst researchers, although it is argued effect sizes should be interpreted with respect to empirical benchmarks relevant to the intervention (Hill, Bloom, Black, & Lipsey, 2006). (Hattie, 2009) argued that only effect sizes greater than .40 are considered to have an impact on pupil outcomes. In Hattie's (2009) synthesis CL demonstrated an effect size of 0.59 compared to individualistic learning. Although it is argued that even if an educational intervention could make small and inexpensive change by an effect size of even as little as 0.1 then this could be a noteworthy improvement (Glass et al, 1981).

Effect sizes are reported in the current study. Cohen's d is reported following t-tests and ANOVA; and Pearson's r following non-parametric tests a summary of the size criteria for each one is presented in Table 3.1.

Table 3. 1 Effect Sizes and Descriptors for Cohens d and Pearson's r.

Size	Cohen's d	Pearson's r
Small	.2	.1
Medium	.5	.3
Large	.8	.5

Adapted from Pallant (2010)

3.4 Power

A power of at least 0.8 at an alpha level of 5% is considered the convention (Wright, 2003), signifying an 80% chance of finding an effect where one exists. A priori power analysis using G power software (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a total sample of 128 participants was needed to achieve an 80% chance of identifying a medium effect (d=.05).

This number of participants was entered in the analyses for the PA measure (n=157) when the schools were combined, meaning that the study had sufficient power to detect an effect if one existed. However, this number of participants for the analyses of the empathy and experience of bullying measure fell marginally below that indicated meaning that the study may not have sufficient power to detect an effect.

Post Host analyses found that for the empathy questionnaire across the whole sample (adapted from the Emotional Literacy Scales (Faupel, 2003)) power was 0.68 (n=97). For the bullying measure (measured by the My Life in Schools questionnaire (Arora & Thompson, 1987) across the whole sample power was 0.78.

When analysing individual school's power was low. Post hoc analysis indicated that the power level for PA, measured by the SIS (Fredrickson & Furnham, 1998) ranged between 0.39 and 0.43. Power for Empathy (adapted from the Emotional Literacy Scales (Faupel, 2003) ranged between 0.29 and .024 and for Bullying (measured by the My Life in Schools questionnaire (Arora & Thompson, 1987) this ranged between 0.32 and .037.

Primary analyses were carried out on whole data sets, combining all schools. Individual schools' analyses are included to provide explore specific school trends which can be triangulated with intervention implementation data. Given the above low power calculations effect sizes will be reported in addition to *p* values, as *p* values are criticised for their limited value (Wright, 2003). Effect sizes are independent of sample size and provide an indication of the magnitude of the differences found (Sullivan & Feinn, 2012).

3.5 PA

PA was measured using the SIS (Fredrickson and Furnham, 1998). The relevant research questions and experimental hypothesis in relation to the variable are presented:

Research Question One: What is the impact of CL on pupils' PA?

Research Question Two: Does the impact of CL on PA differ depending on pupils' initial levels of PA?

3.5.1 Grouping data from multiple schools

Data from three classes across two schools were combined for analyses. It was therefore important to establish the validity of treating them as one data set. A one-way ANOVA indicated that the data from the three classes were not significantly different at pre-test in each context on:

SIS Play Same Gender (PSG) ($F(5, 152) = .737$, $p = .597$)

SIS Work Same Gender (WSG) ($F(5, 152) = 1.54$, $p = .181$)

However, the data from the three classes were statistically significantly different at pre-test on:

SIS Play Different Gender (PDG) ($F(5, 152) = 3.48$, $p = .005$)

SIS Work Different Gender (WDG) ($F(5, 152) = 5.372$, $p = .000$)

Inspection of the means plots and post hoc comparisons using the Tukey HSD test indicated this difference was associated with School B Year 5. Excluding School B Year 5 classes from this variable on the analysis was considered, however it was judged that combining all classes would represent the real-world variances between schools and classes using CL. Interpretations considered this difference and data were analysed based by individual classes.

3.6 Play Same Gender (PSG)

3.6.1 Main analysis

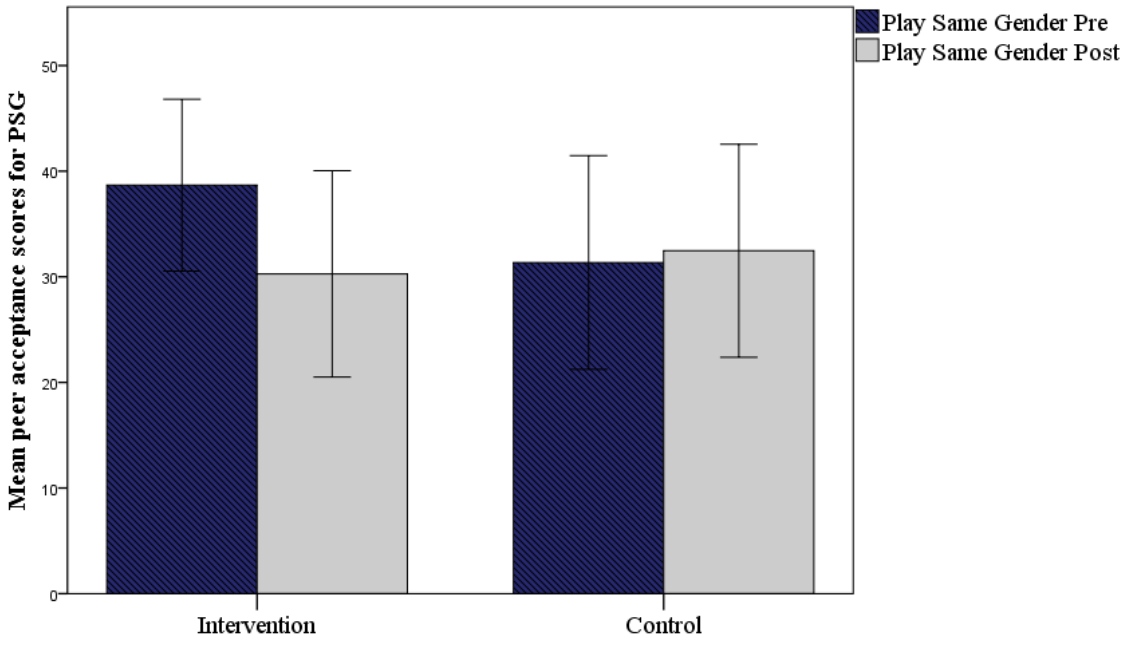
Data from all four schools were combined. Data was not normally distributed, although equality of variance could be assumed

3.6.1.1 Descriptive statistics

Table 3. 2 A table to show the Mean, Standard deviation and Median for PA pre and post-test scores and gain scores for PSG

		n	Mean	SD	Median	Gain Score (Mean)	SD	Gain Score (Median)
Intervention	Pre	79	38.68	36.30	36.36	-8.40	39.96	-9.72
	Post	79	30.27	43.61	33.36			
Control	Pre	78	31.35	44.88	42.26	1.12	40.53	-2.68
	Post	78	32.47	44.74	42.86			

Figure 3. 1 A bar chart to show scores at pre and post-test for the intervention and control group for PSG.



3.6.2 Main statistical analysis

The pre-test means for the intervention group were slightly larger than the Comparison group, although a Mann-Whitney U test indicated that there was not a significant difference between pre-test scores $U = 2917$, $Z = -.574$, $p = .566$. Mean scores increased slightly for the control group. However as illustrated in Figure 3.1 there was a decrease in PA scores in the intervention group, which was not in the direction of the experimental hypothesis. A Mann-Whitney U test indicated that there was not a significant difference in gain scores for intervention pupils (Md=-9.72. n=79) and control pupils ($Md = 2.68$, n = 78), $U = 27645$, $Z = -1.11$, $p = .266$). This represented a small effect ($r = -.09$).

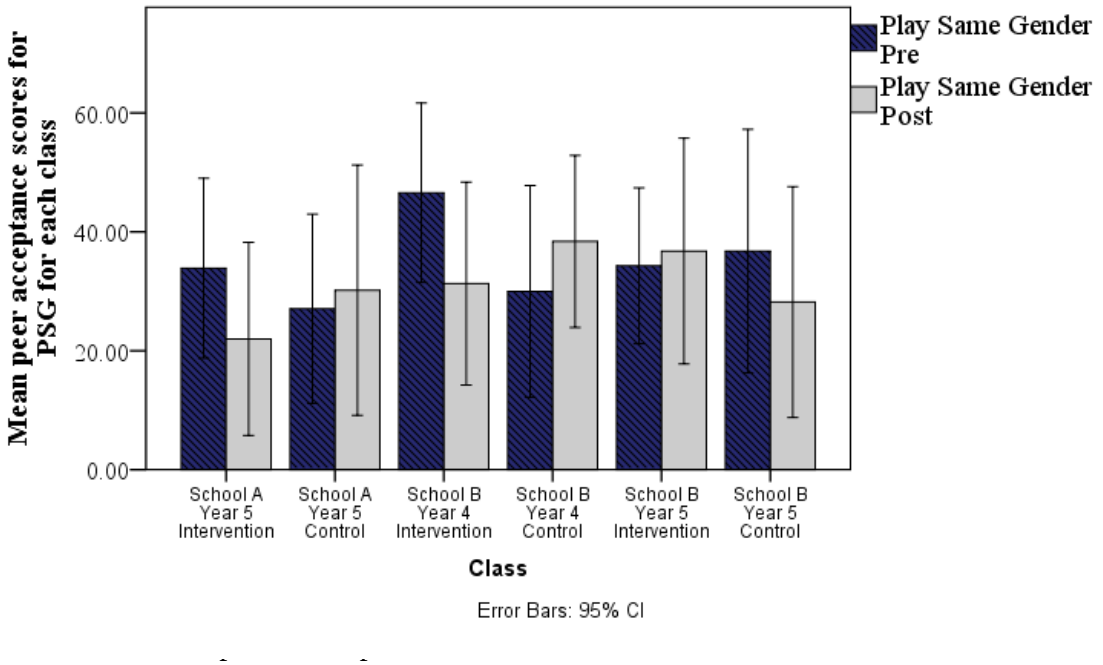
3.6.3 Analysis by individual class groups

3.6.3.1 Descriptive statistics

Table 3. 3 A table to show the Mean, Standard deviation and median for PA pre and post-test scores and gain scores in each class for PSG

Class	Intervention		n	Mean	SD	Median	Gain score	SD	Gain score (median)
School A Year 5	Intervention	Pre	24	33.89	35.84	35.42	-11.90	41.37	-25.90
		Post	24	21.99	38.46	21.11			
	Control	Pre	24	27.08	37.65	37.50	3.10	42.13	7.94
		Post	24	30.19	49.88	40.18			
School B Year 4	Intervention	Pre	29	46.57	39.65	60.00	-15.26	35.70	-12.50
		Post	29	31.31	44.87	37.50			
	Control	Pre	28	29.99	45.90	43.65	8.41	38.70	2.68
		Post	28	38.40	37.32	50.00			
School B Year 5	Intervention	Pre	26	34.30	32.47	30.33	2.47	42.32	14.32
		Post	26	36.76	46.98	47.25			
	Control	Pre	26	36.75	50.73	61.25	-8.56	40.61	-15.34
		Post	26	28.19	48.05	44.16			

Figure 3. 2 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for PSG..



was not in the direction of the experimental hypothesis.

3.6.3.2 Statistical analysis for individual class groups

Individual Data from each class was scrutinised to explore the impact at a class level.

Data across all schools was not normally distributed. The Mann-Whitney U test was therefore used to establish whether there was a significant difference in gain scores for pupils in the CL group and control group. A Wilcoxon Signed Rank test was used to establish whether there was a significant difference in PSG scores from pre and to post.

Table 3. 4 A Table to show results from statistical analysis for each class for PA for PSG.

Class	Results from statistical analysis
School A Year 5	Wilcoxon Signed Ranks Test: No statistically significant differences from to pre to post-test in either the intervention (Z = -1.380, p = .168) or control group (Z = -.584, p = .559) with a medium (r = - .3) and small effect size (r = - .2) respectively. Mann-Whitney U test: No statistically significant difference in gain scores for intervention pupils (Md = -25.90, n = 24) and control pupils (Md = 7.94, n = 48), U = 217.5, Z = - 1.45, p =.146). This represented a small effect (r = -.2).
School B Year 4	Wilcoxon Signed Ranks Test : No statistically differences from pre to post test in either the intervention (Z = -1.890), p = .059) or control group (Z = -.673, p = .501) with a medium (intervention r = .4) and small (control r = .1) effect size. Mann-Whitney U test: No statistically significant difference in scores for the intervention pupils (Md = -12.50, n = 29) and control group (Md = 2.68) , U = 287.5, Z = -1.892, p = .058 representing a medium effect (r = - .3).
School B Year 5	Wilcoxon Signed Ranks Test revealed no statistically differences from pre to post test in either the intervention (Z = -.323, p = .747) or control group (Z = -.1932, p = .053) with a very small effects (intervention r = - .06; control r = - .4) effects. Mann-Whitney U test revealed there was no statistically significant change in scores for the intervention pupils (Md= 14.32, n=26) and control group (Md = -15.33), U = 287.5, Z = -1.41, p = .159 representing a small effect (r = -.20).

3.7 Play different gender (PDG)

3.7.1 Main analysis

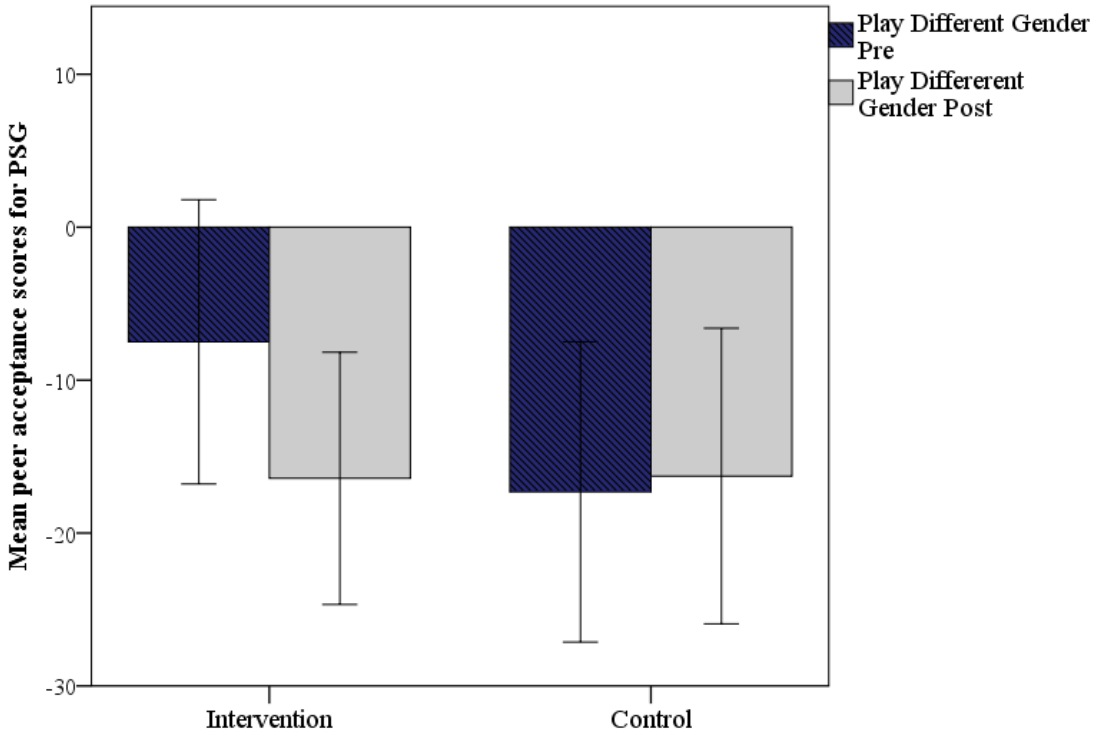
Tests of normality indicated data was not normally distributed, although equality of variance could be assumed.

3.7.1.1 Descriptive statistics

Table 3. 5 A table to show the Mean, Standard deviation and Median for PA pre and post-test scores and gain scores for PDG.

		N	Mean	SD	Median	Gain score (mean)	SD	Gain score (Median)
Intervention	Pre	79	-7.50	41.50	-12.50	-8.93	37.46	-5.00
	Post	79	-16.42	36.84	-20.00			
Control	Pre	78	-17.31	43.56	-26.79	1.03	44.3	6.60
	Post	78	-16.28	42.85	-21.11			

Figure 3. 3 A bar chart to show scores at pre and post-test for the intervention and control group for PDG.



3.7.1.2 Main statistical analysis

Pre-test means for the control group were slightly larger than the intervention group, however, a Mann-Whitney U test indicated that there was not a significant difference between pre-test scores (U = 2593, Z = -1.712, p = .087). Mean scores decreased for the intervention group from pre-test to post-test and decreased slightly for the control group. A Mann-Whitney U test indicated that there was not a significant difference in gain scores for intervention pupils (Md = -5.00, n =79) and control pupils (Md = 6.60, n =78), (U = 2523, Z = -1.96, p = .50). This represented a small effect (r = .2).

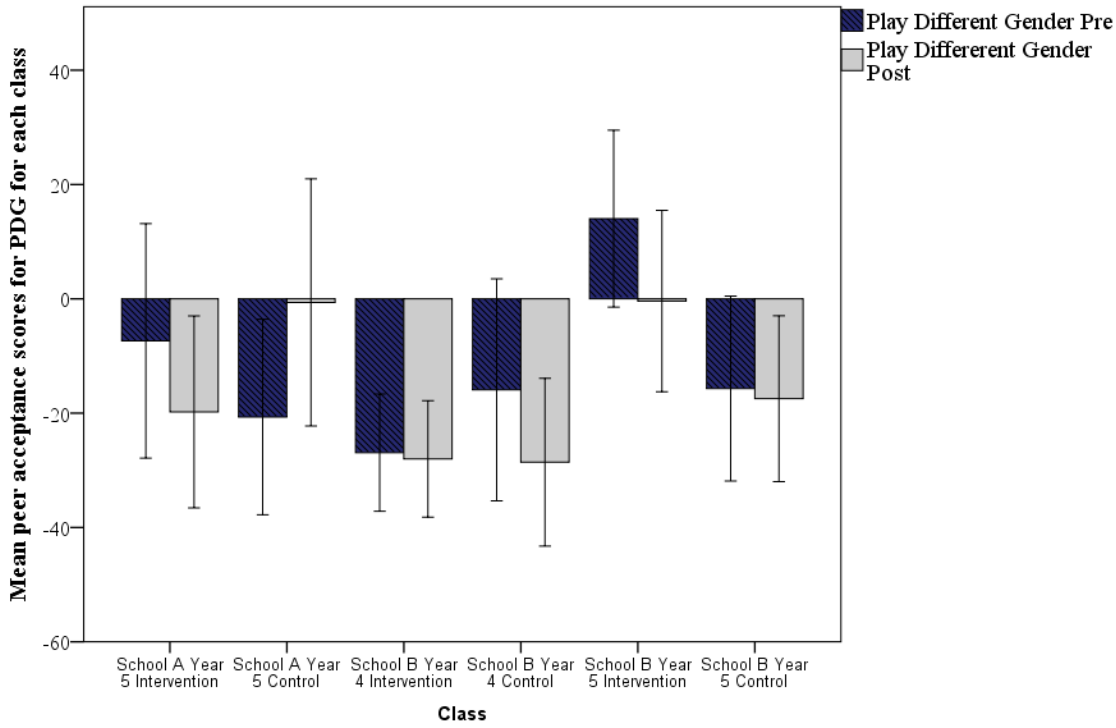
3.7.2 Analysis by individual classes groups

3.7.2.1 Descriptive statistics

Table 3. 6 A table to show the Mean, Standard deviation and median for PA pre and post-test scores and gain scores in each class for PDG

Class	Intervention		N	Mean	SD	Median	Gain Score (mean)	SD	Gain Score (median)
School A Year 5	Intervention	Pre	24	-7.36	48.58	-10.00	-12.41	53.86	-.550
		Post	24	-19.78	39.76	-20.00			
	Control	Pre	24	-20.68	40.52	-16.67	20.04	-40.52	30.07
		Post	24	-0.636	51.20	-6.25			
School B Year 4	Intervention	Pre	29	-26.90	26.99	-33.33	-1.11	25.93	-5.56
		Post	29	-28.01	26.76	-27.27			
	Control	Pre	28	-15.92	50.11	-30.95	-12.66	38.88	-10.42
		Post	28	-28.65	37.86	-25.00			
School B Year 5	Intervention	Pre	27	15.19	38.05	012.50	-14.42	29.04	-8.41
		Post	26	-0.39	39.28	0.00			
	Control	Pre	26	-15.69	40.06	-27.76	-1.77	40.45	8.58
		Post	26	-17.46	35.92	-18.05			

Figure 3. 4 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for PDG.



Ass Illustrated in Figure 3.4 the mean score decreased across all intervention groups with increases in the control group in School A Year 5 control.

3.7.2.2 Statistical analysis for individual classes

Individual Data from each class was scrutinised to explore the impact at a class level. Data from School A Year 5 and School B Year 5 were normally distributed, and an independent samples t-test was used to test for significance. Data from School B Year 4 was not normally distributed therefore the Man Whitney U t\and Wilcoxon Signed Rank test was used.

Table 3. 7 A Table to show results from statistical analysis for each class for PA for PSG

Class	Results
School A Year 5	Statistically significant difference in gain scores for the intervention (*M* = 12.41, SD = 53.87) and control group (*M* = 20.04, SD = 49.01), t(46) = -2.18, *p* = .034) with no significant differences between the groups at pre-test, t(46) = .641, *p* = .525 representing a medium effect (*d* = -.6).
School B Year 4	Wilcoxon Signed Rank: revealed no significant differences from pre to post-test in either the intervention (*Z* = .592, *p* =.554) or control group (*Z* = -1.562, *p* = .118) representing a small (intervention *r* = .1) and medium effect (control r= - .3). Mann-Whitney U: No statistically significant difference in gain scores for the intervention pupils (*Md* = -5.56) and control pupils (*Md* = -10.42), (*U* = 359, *Z* = -.750, *p* = .453) representing a very small effect size (*r* = .09).
School B Year 5	Statistically significant difference in gain scores for the intervention (*M* = -14.43, SD = 29.04) and control group (*M* = -1.78, SD = 7.93), *t*(50) = -1.30, *p* = .201, representing a small but approaching medium effect size (*d* = -.4).

3.8 Work same gender (WSG)

3.8.1 Main analysis

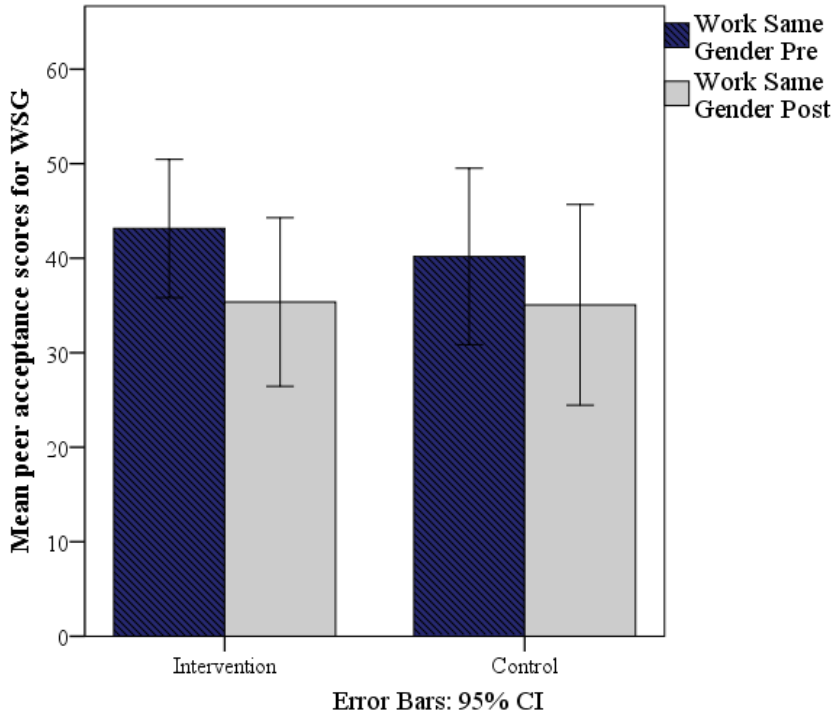
Tests of normality indicated data deviated significantly from the norm although equality of variance could be assumed.

3.8.1.1 Descriptive statistics

Table 3. 8 A table to show the Mean, Standard deviation and Median for PA pre and post-test scores and gain scores for WSG

		N	Mean	SD	Median	Gain score (mean)	SD	Gain score (median)
Intervention	Pre	80	43.14	32.70	45.45	-7.78	33.54	-5.00
	Post	79	35.36	39.84	44.44			
Control	Pre	78	40.18	41.39	55.05	-5.12	32.45	-5.12
	Post	78	35.06	47.08	50.00			

Figure 3. 5 A bar chart to show scores at pre and post-test with confidence intervals for the intervention and control group for WSG.



3.8.1.2 Main Statistical Analysis

The pre-test means for the intervention group was slightly larger than the control group. Mean scores decreased slightly for both groups from pre-test to post-test. A Mann-Whitney U test indicated that there was not a significant difference between pre-test scores U = 3053, Z = -.100, p =.920 and no significant difference in gain scores for intervention pupils (Md = -5.00, n =79) and control pupils (Md = 5.12 n =78), U = 2940, Z = -.495, p =.621). This represented a very small effect (r = -.04).

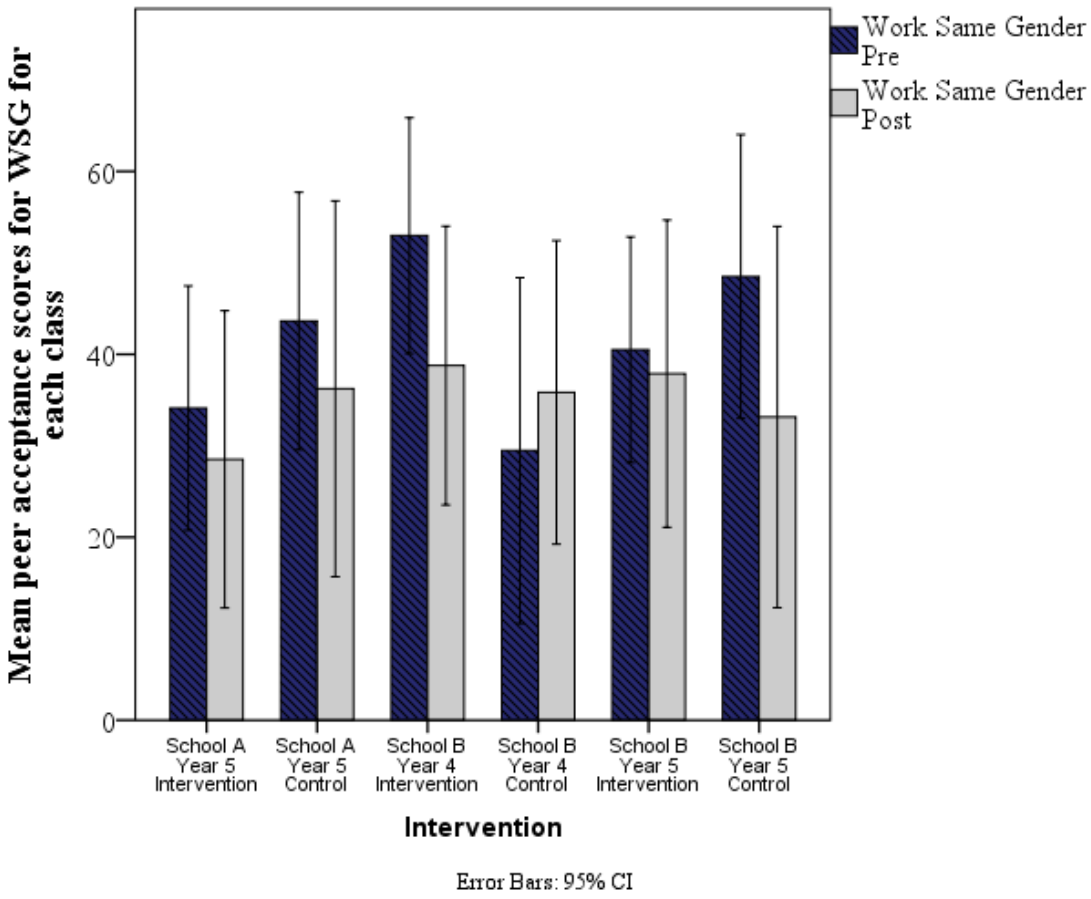
3.8.2 Analysis by individual class groups

3.8.2.1 Descriptive Statistics

Table 3. 9 A table to show the Mean, Standard deviation and median for PA pre and post-test scores and gain scores in each class for WSG.

Class	Intervention		N	Mean	SD	Median	Gain Score	SD	Gain score (Median)
School A Year 5	Intervention	Pre	24	35.13	31.63	41.43	-5.61	36.07	.00
		Post	24	28.52	38.47	35.42			
	Control	Pre	24	43.64	33.37	53.57	-7.41	35.20	-3.47
		Post	24	36.23	48.66	43.75			
School B Year 4	Intervention	Pre	29	52.97	33.89	60.00	-14.20	27.70	-11.11
		Post	29	38.77	40.05	45.45			
	Control	Pre	28	29.48	48.74	40.18	6.36	27.99	3.87
		Post	28	35.84	42.82	40.18			
School B Year 5	Intervention	Pre	26	42.70	32.05	42.22	-2.61	39.36	2.09
		Post	26	37.88	41.82	47.73			
	Control	Pre	26	48.51	38.41	63.64	-15.36	31.55	-11.61
		Post	26	33.14	51.60	52.28			

Figure 3. 6 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for WSG.



3.8.2.2 Statistical Analysis for individual class groups

Individual Data from each class was scrutinised to explore the impact at a class level. Data across all schools were not normally distributed therefore the Man Whitney U t\and Wilcoxon Signed Rank test was used.

Table 3. 10 A Table to show results from statistical analysis for each school and class for WSG.

School		Results from statistical analyses
School Year 5	A	Wilcoxon Signed Rank test: No statistically significant differences from pre to post-test in either the intervention (Z = -1.38, p = .168) or control group (Z = -.584, p = .559) with a very small (r=.03) and small effects (r=.1) respectively. Mann-Whitney U test: No statistically significant difference in gain scores for intervention pupils (Md = -.00, n= 24) and control pupils (Md = -3.47, n= 24), (U=279, Z = -.186, p = .853). This represented a very small effect (r = .02).
School Year 4	B	Wilcoxon Signed Rank: Statistically significant differences from pre to post-test in the intervention group (Z = -2.72, p = .007) but not in the control group **(**Z**= -** .922, p = .321) with a large (r = .5) and small effect size (r= .2) respectively. Mann- Whitney U test: No statistically significant difference in gain scores for intervention pupils (Md = -11.11, n =28) and control pupils (Md = 3.87, n =29), (U = 250.5 Z = -2.48, p = .013). This represented a medium effect (r = .3)
School Year 5	B	Wilcoxon Signed Rank: No statistically significant differences from pre to post-test in the intervention (Z= -.188, p = .851) group but there were statistically significant differences in the control (Z = -2.26, p = .024) with a very small (r = -.03) and medium effect size (r= -.4) respectively. Mann-Whitney U test: No statistically significant difference in gain scores for the intervention (Md = 2.09, n =26) and control group (Md = -11.61), n =29), U = 266.5, Z = -1.31, p = .191). This represented a small effect (r =.2).

3.9 Work different gender (WDG)

3.9.1 Main analysis

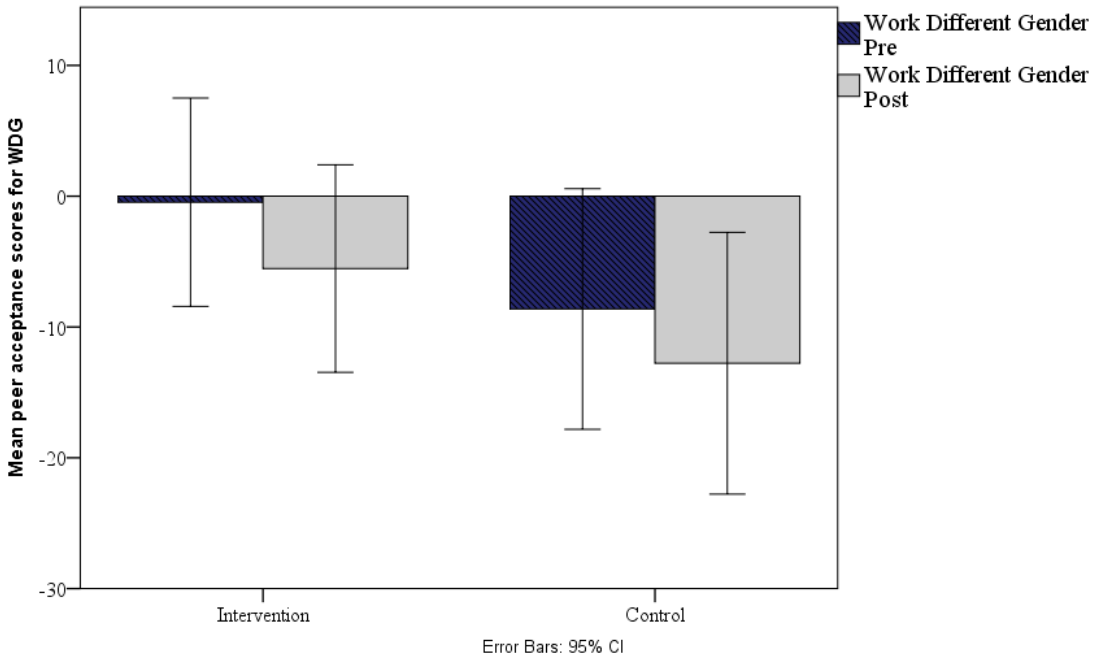
Tests of normality indicated that data was normally distributed although there was significant variance between the groups so the 'Equality not assumed' statistic will be reported were appropriate.

3.9.1.1 Descriptive statistics

Table 3. 11 A table to show the Mean, Standard deviation for PA pre and post-test scores and gain scores for WDG

		N	Mean	SD	Gain score	SD
Intervention	Pre	79	-0.46	35.57	-5.07	33.02
	Post	79	-5.53	35.42		
Control	Pre	78	-8.62	40.79	-4.14	27.97
	Post	78	-12.77	44.37		

Figure 3. 7 A bar chart to show scores at pre and post-test with confidence intervals for the intervention and control group for PSG.



3.9.1.2 Main statistical analysis

The pre-test means for the control group were slightly larger than the intervention group, however an independent samples t-test indicated that differences between the two groups at pre-test were not statistically significant, $t(155) = .337$, $p = .183$. Mean scores decreased for both groups from pre to post-test but there was no statistical significance different in gain scores for intervention pupils ($M = -5.07$, SD 33.02) and control pupils ($M = -4.14$, SD =27.97); $t(155) = -.189$, $p = .850$ representing a very small effect ($d = .03$).

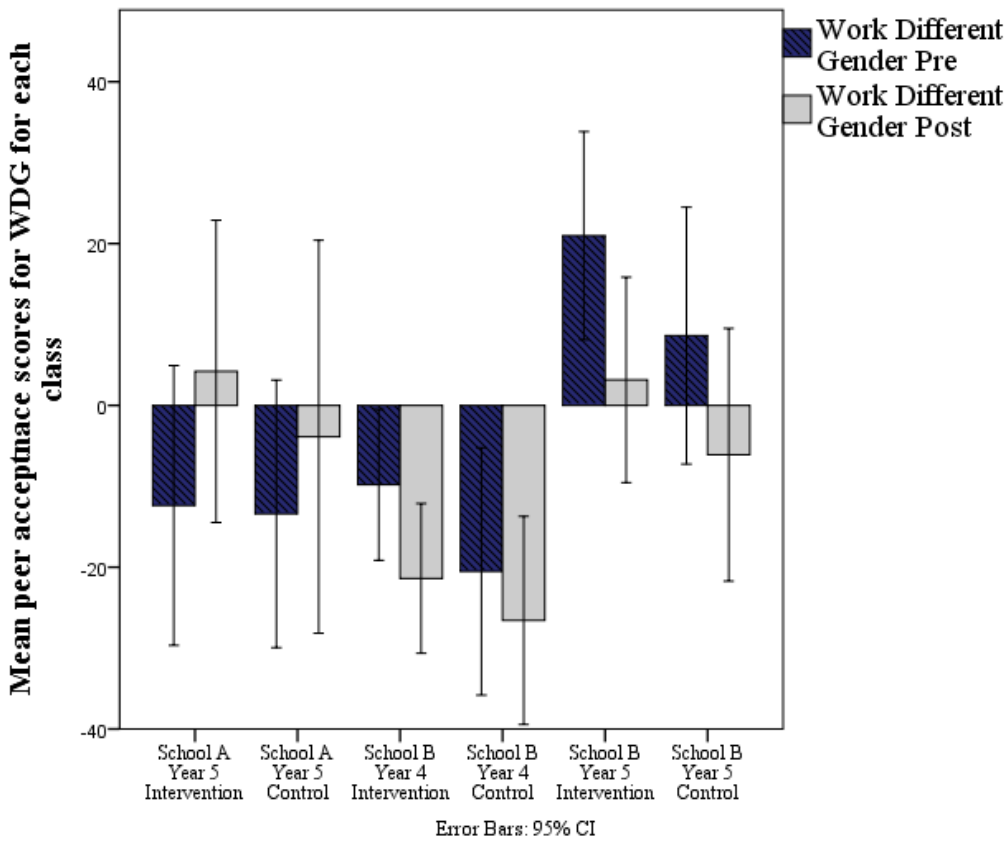
3.10 Analysis with individual classes

3.10.1 Descriptive statistics

Table 3. 12 A table to show the Mean, Standard deviation and median for PA pre and post-test scores and gain scores in each class for WDG.

Class	Intervention		N	Mean	SD	Median	Gain Score	SD	Gain score (median)
School A Year 5	Cooperative Learning	Pre	24	-12.37	40.94	-10.00	16.59	37.65	23.31
		Post	24	4.21	44.24	11.69			
	Comparison	Pre	24	-13.41	39.16	-17,36	9.54	35.38	7.64
		Post	-3.88	57.53	-3.88	-11.11			
School B Year 4	Cooperative Learning	Pre	29	-9.82	24.53	-10.00	-11.57	24.32	-11.11
		Post	29	-21.38	24.32	-22.22			
	Comparison	Pre	28	-20.53	39.44	-23.61	-6.06	25.65	-1.79
		Post	28	-26.53	33.17	-28.57			
School B Year 5	Cooperative Learning	Pre	26	22.05	31.75	20.20	-17.82	27.18	-16.67
		Post	26	3.15	31.45	11.53			
	Comparison	Pre	26	8.62	39.29	31.87	-14.72	15.73	27.18
		Post	26	-6.10	38.67	31.45			

Figure 3. 8 A bar chart to show overall mean scores for PA for each class at pre and post-test, for the intervention and control group for WDG.



As illustrated Figure 3.8 in school A Year 5 there was an increase in scores from pre to post-test in both the intervention and control group. In both School B year 4 and 5 classes mean scores decreased in both intervention and control group.

3.10.2 Statistical analysis for individual class groups

Data from each class was explored. Data from School A Year 5 and School B Year 4 was normally distributed therefore an independent sample t-test was A Mann-Whitney U and Wilcoxon Signed Rank tests were used to test for significance in School B Year 5.

Table 3. 13 A Table to show results from statistical analysis for each class for PA for WDG

School	Results from statistical analyses
School A Year 5	There was no statistically significant difference in gain scores for the intervention (*M* = 16.59, SD = 37.65) and control group (*M* = 9.54, SD = 35.38); t(46) = .669, *p* = .507) with no significant differences between the groups at pre-test, t(46) = .090, *p* = .929 representing a small effect (*d* = .2).
School B Year 4	There was no statistically significant difference in gain scores for the intervention (*M* = -11.57, SD = 24.84) and control group (*M* = - 6.06, SD =25.65); *t*(55) = -.824, *p* = .414) with no significant differences between the groups at pre-test.; *t*(46) = 1.24, *p* = .222, representing a small effect (*d* = .2).
School B Year 5	Wilcoxon Signed Rank test: Statistically significant differences from pre to post-test the intervention (*Z* = -2.960, *p* = .003) and control group (*Z* = -3.53, *p* = .000) with a large effects (intervention *r* =- .6; control *r* = -.7). Mann-Whitney U test: No statistically significant difference in gain scores for intervention pupils (*Md* = -16.67, *n* =26) and control pupils (*Md* = 27.18, *n* =24), *U* = 305.5, *Z* = -.595, *p* = .552), representing a very small effect *r* = -.08).

3.10.3 PA: Summary of findings

When combining data from all schools, there were no statistically significant differences in gain scores between the intervention and control group. Due to significant pre-test differences associated with School B Year 5 an analysis was conducted with this class removed. With School B removed there was a statistically significant difference between gain scores between the intervention (*Md* = -12.50, *n*= 53) and control group (*Md* = 5.36, *n* =52); (*U*= 1001.50, *Z* = -2.414, *p* = .016) in the PSG context representing a small effect (*r* = - .2).

In terms of individual school level changes, findings were predominately in favour of the intervention classes, however, this was associated with a decrease in PA scores for the intervention group:

PSG: No statistically significant differences in any of the three classes.

PDG: Statistically significant differences in School A Year 5 (*p* = .034) representing a medium effect (*d*=.6) However, this was a statistically significant decrease in the intervention group

WSG: Statistically significant differences between gain scores in School B Year 4 (*p* = .013) with a medium effect (*r* = .3). However, this was a statistically significant decrease between pre and post-test scores in the intervention group (*p* =.007) with a medium effect (*r* =.5).

WDG: Statistically significant differences from pre to post-test the intervention group (*p* = .003) and control group (*p* =.000) with a medium effect (intervention *r* = -.6; control *r* = -.7). However, this was a statistically significant decrease in both groups.

3.11 PA: A differential impact

Pupils data was divided into three groups at the analysis stage according to their initial levels of PA. Groups were calculated using the interquartile range:

Low **below 25th percentile**

Medium **at or above 25th to at or below the 75th percentile**

High **above 75th percentile**

Analysis compared changes in scores for the pupils with low, medium and high levels at the start of the intervention to explore the impact of CL The analysis aimed to explore whether CL leads to gains or losses for particular groups which may not be visible within the main analysis using overall means.

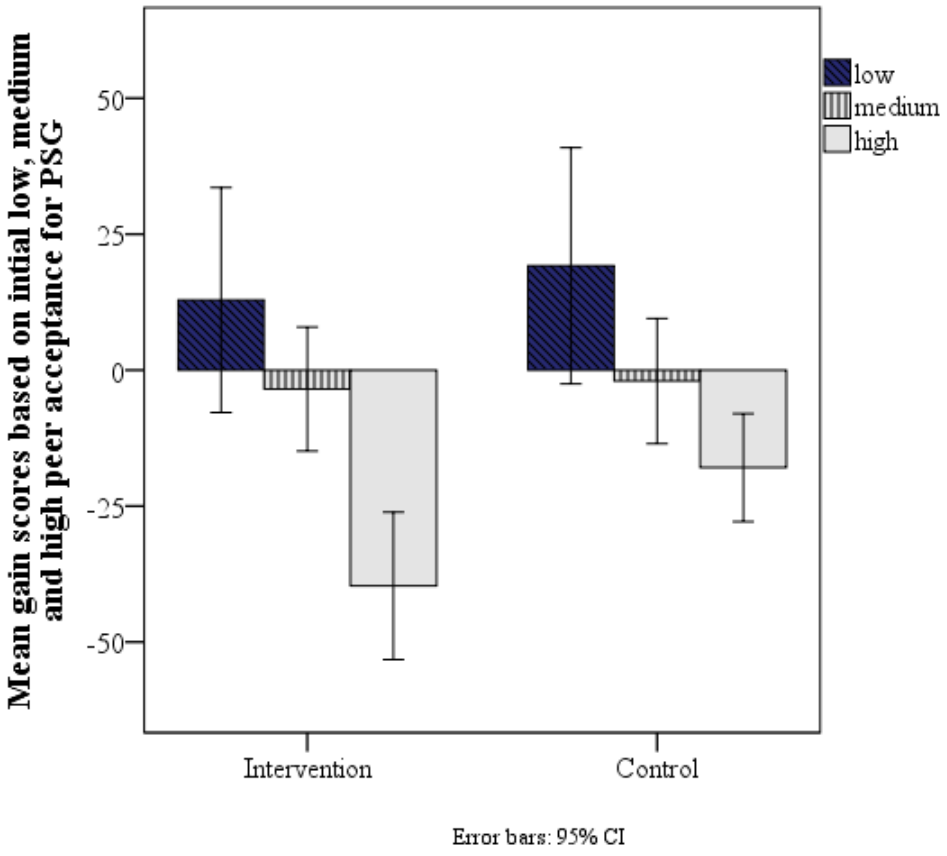
It is acknowledged that comparing high and low groups could increase the risks associated with regression to the mean, whereby it is most likely for low scores to increase and high scores to increase from pre to post-test, despite the impact of an intervention. Therefore, results need to be interpreted with caution.

3.11.1 Peer same gender

Table 3. 14 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in Play Same Gender (PSG) context.

	Initial PA level	N	Gain score (mean)	SD	Gain Score (median)
Intervention	Low	16	12.89	38.82	21.67
	Medium	45	-3.46	37.95	0.00
	High	18	-39.68	27.27	-37.63
Control	Low	25	19.21	52.63	0.00
	Medium	35	-2.01	33.53	5.36
	High	18	-17.93	19.95	-21.31

Figure 3. 9 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for Play Same Gender (PSG) context.



3.11.2 Statistical analysis

As data was not normally distributed the parametric alternative to the one way between-groups analysis of variance (Kruskal-Wallis) was used to explore the impact of cooperative learning on gain scores for the Play Same Gender (PSG) context for children with initial low, medium or high PA levels. A Kruskal-Wallis revealed a statistically significant difference in gain scores across the three levels of PA in the intervention group x^2 (2, n=79) = 16.89, p =.000. Children with initially low levels of PA recorded a positive gain score (*Md* = -21.67) compared to those with the high levels of PA who recorded a higher negative score (*Md* = -37.63). There was also a statistically significant difference in gain scores across the three levels of PA in the Control group x^2 (2, n=78) =8.70, p=.013.

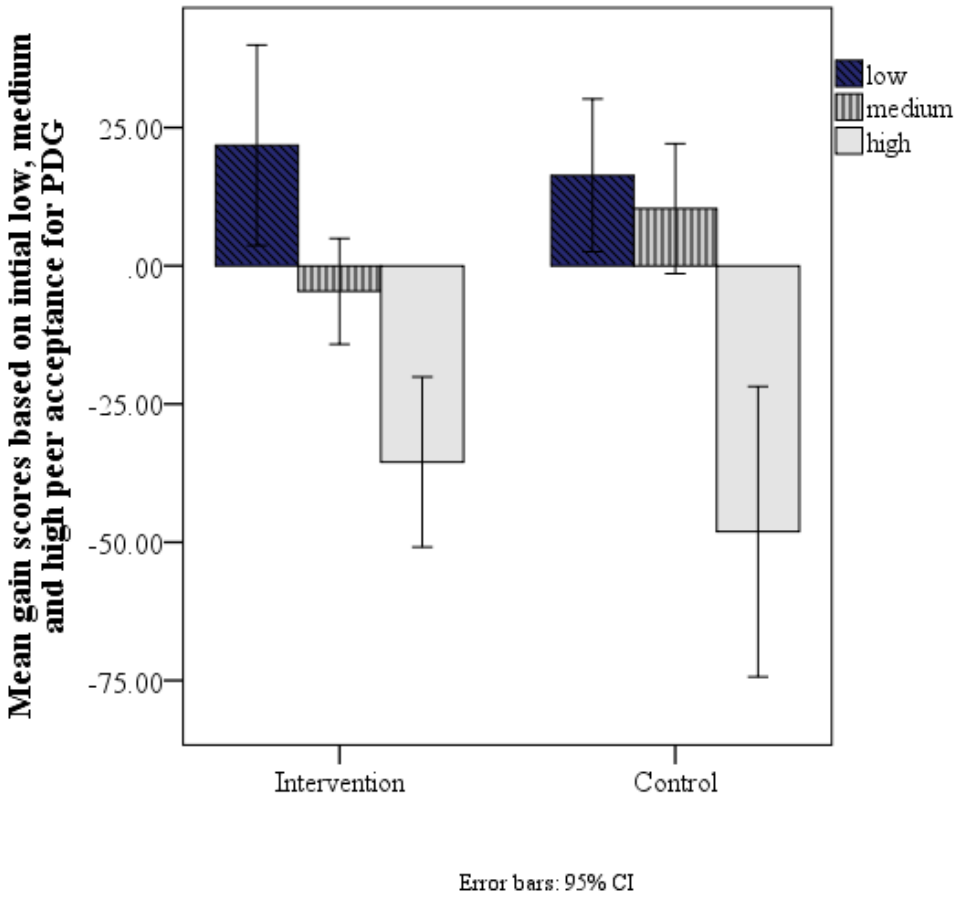
Man Whitney U tests indicated significant differences were between low and high groups in the intervention (*Z*=-3.72, *p* =.000, *r* = -.06) and control groups (*Z = 287.5*, *p*=.008, *r* = -.40) and medium and high groups in the intervention (*Z = -3.416, p* =.001, r =.4) and control group (*Z =-2.37*, *p* = .018, r = -.3) levels of PA. Children with initially low levels of PA recorded a positive mean gain score and those with initially high levels of PA recording a negative mean gain score. There is a similar pattern of a negative change in both conditions for children with initially high levels of PA with a greater decrease in scores in the intervention condition.

3.12 Play different gender

Table 3. 15 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in PDG context.

	Initial PA level	N	Gain score (mean)	SD	Gain Score (median)
Intervention	Low	14	26.79	31.42	16.17
	Medium	42	-4.62	30.68	-4.29
	High	23	-35.50	35.57	-40.00
Control	Low	25	16.36	33.49	11.11
	Medium	38	10.34	35.80	13.60
	High	15	-48.09	47.39	-50.89

Figure 3. 10 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for PDG.



3.12.1 Statistical analysis

The Kruskal-Wallis test revealed a statistically significant difference in gain scores across the three levels of PA in the intervention group x^2 (2, n=79) = 15.55, p=.000. Children with initially low levels of PA recorded a positive gain score (Md=26.79) compared to those with the high levels of PA who recorded a higher negative score (Md= -35.50)

There was also a statistically significant difference in gain scores across the three levels of PA in the Comparison group x^2 (2, n=79) = 16.40, *p*=.000. Follow up Man Whitney U tests indicated significant differences were between low and medium group in the intervention group (*Z = -2.61, p* = .009, *r* = -.3), between low and high in both the intervention (*Z = -3.85, p* = .000, r=-.6) and control (*Z = -3.689, p* = .000, *r* = -.6)

and between medium and high in both intervention ($Z = -3.06$, $p = .000$, $r = -.4$) and control ($Z = -3.69$, $p = .000$, $r = -.5$) groups.

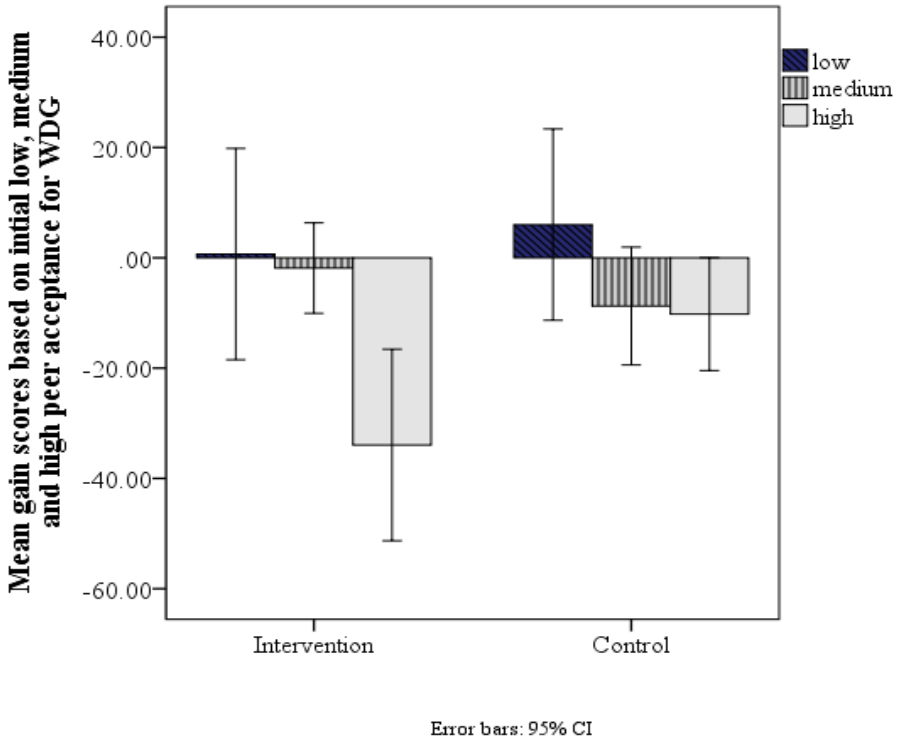
Children with initially low levels of PA recorded a positive mean gain score and those with initially high levels of PA recorded a negative mean change. In the intervention group, children with initial low PA levels had a higher positive gain score compared to the control group.

3.13 Work same gender (WSG)

Table 3. 16 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in WSG.

	Initial PA level	N	Gain score (Mean)	SD	Gain Score (Median)
Intervention	Low	18	0.69	38.52	0.00
	Medium	45	-1.86	27.31	0.00
	High	16	-33.94	32.55	-27.27
Control	Low	21	6.00	38.11	-7.78
	Medium	39	-8.77	32.94	-6.94
	High	18	-10.20	20.59	-11.61

Figure 3. 11 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for Work Same Gender (WSG) context.



3.13.1 Statistical analysis

The Kruskal-Wallis test revealed a statistically significant difference in gain scores across the three levels of PA in the intervention group. x^2 (2, n=79) = 10.98, p = .004 Children with initially high levels of PA recorded a high negative gain score (Md = -27.27) compared to those with low or medium levels of PA. Follow up Man Whitney U tests indicated significant differences were between low and high groups in the intervention group (Z = *-2.49*, p = .012, r = -.4) and medium and high groups in the intervention group (Z = *3-33*, p = .001, r = -.4).

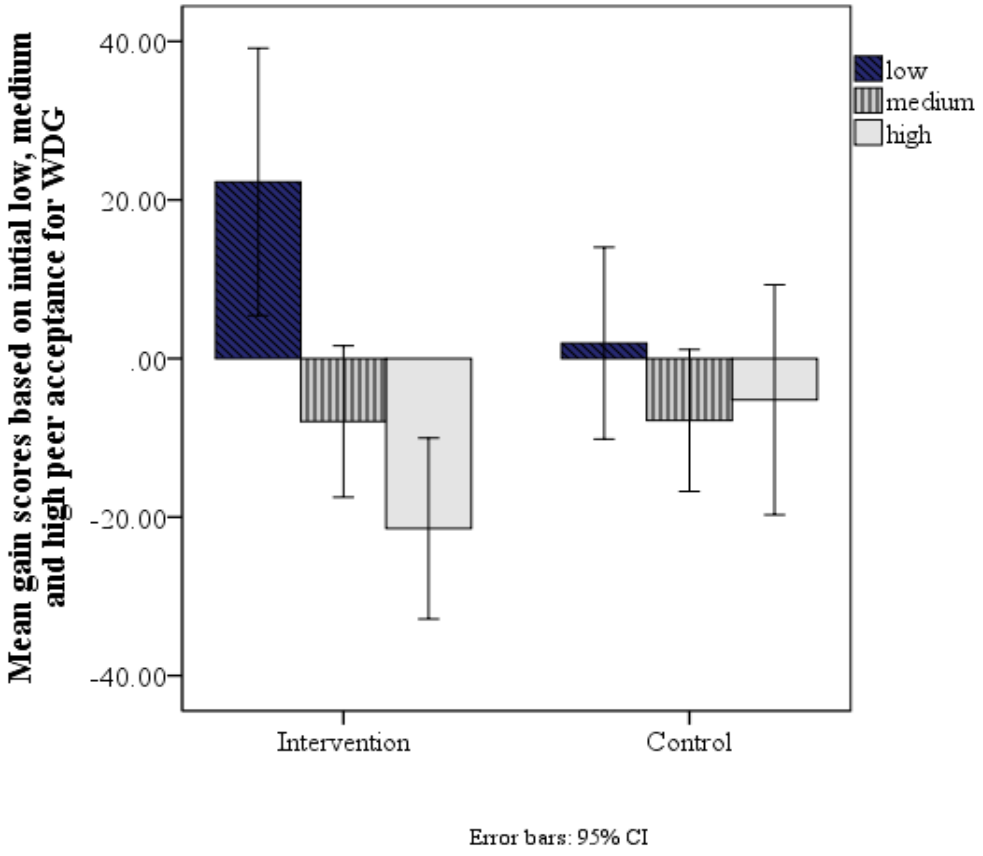
There were no statistically significant difference in gain scores across the three levels of PA in the Comparison group x^2 (2, n=79) = 4.120, p = .127.

3.14 Work different gender

Table 3. 17 A table to show mean gain scores for pupils based on initial low, medium and high levels of PA in the intervention and control conditions in WSG.

	Initial PA level	N	Gain Score (mean)	SD	Gain score (median)
\Intervention	Low	16	22.26	31.63	16.67
	Medium	44	-7.95	31.45	-19.23
	High	19	-21.44	23.69	-19.23
Control	Low	24	1.94	28.62	0.00
	Medium	34	-7.81	25.68	-11.11
	High	20	-5.22	30.99	-10.61

Figure 3. 12 A graph to show mean gain scores for pupils based on initial low, medium and high PA scores for Word Different Gender (WDG) context.



3.14.1 Statistical analysis

The Kruskal-Wallis test revealed a statistically significant difference in gain scores across the three levels of PA in the intervention group x^2 (2, *n*=79) = 15.22, *p* = .000. Children with initially low levels of PA recorded a positive gain (*Md* = 16.67) and children with initially high levels of PA recorded a high negative gain score (*Md* = -19.23). Follow up Man Whitney U tests indicated significant differences were between low and medium group (*Z = -2.99, p* = .003, *r*=-.4) and low and high groups (*Z = -3.81*, p = .000, *r* = -.6).

There was no statistically significant difference in gain scores across the three levels of PA in the Comparison group x^2 (2, n=79) = 4.120 p=.445

3.15 Empathy

Research Question 3: What is the impact of CL on pupils' empathy

Experimental Hypothesis: There will be a statistically significant increase in pupil's self-reported empathy following CL when compared to pupils in the control condition.

As discussed in Section 2.7.3.1 the empathy measure used was adapted from the Emotional Literacy Scales pupil checklist (Faupel, 2003). Due to low reliability of the empathy dimension within the scale, despite the overall reliability being high (0.76), the internal consistency of the measure was re-calculated. The reliability of a scale can vary dependent on the sample. The Cronbach's alpha coefficient of a scale should ideally be above .7 (De Vellis, 2003). In the current study, the Cronbach alpha coefficient was .4. Due to low reliability of the measure, only descriptive data is reported.

3.15.1 Descriptive Statistics

Table 3. 18 A table to show Mean, Standard deviation and gain scores pre-test and post-test on total empathy scores.

		N	Mean	SD	Gain score	SD
Intervention	Pre	50	16.12	2.69	- 0.26	2.72
	Post	50	15.90	2.43		
Control	Pre	47	16.26	2.69	- 0.66	2.31
	Post	47	15.60	2.63		

Figure 3. 13 A bar chart to show mean pre and post-test total empathy scores for intervention and control group.

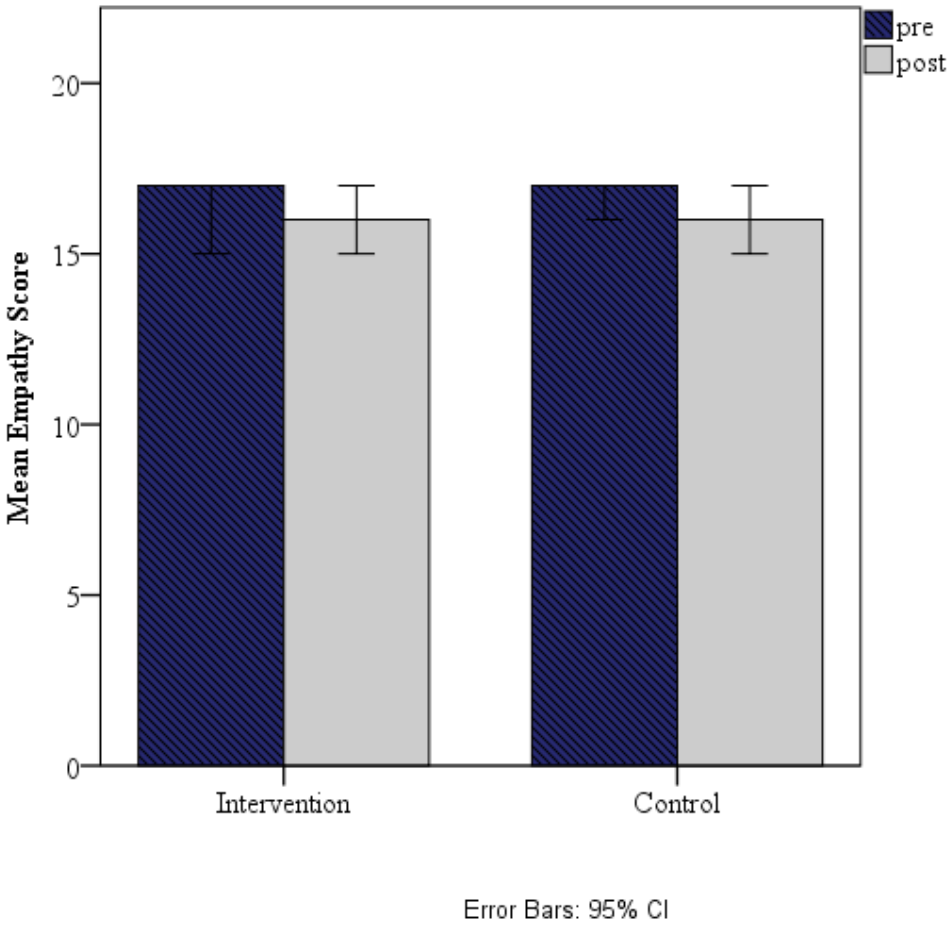
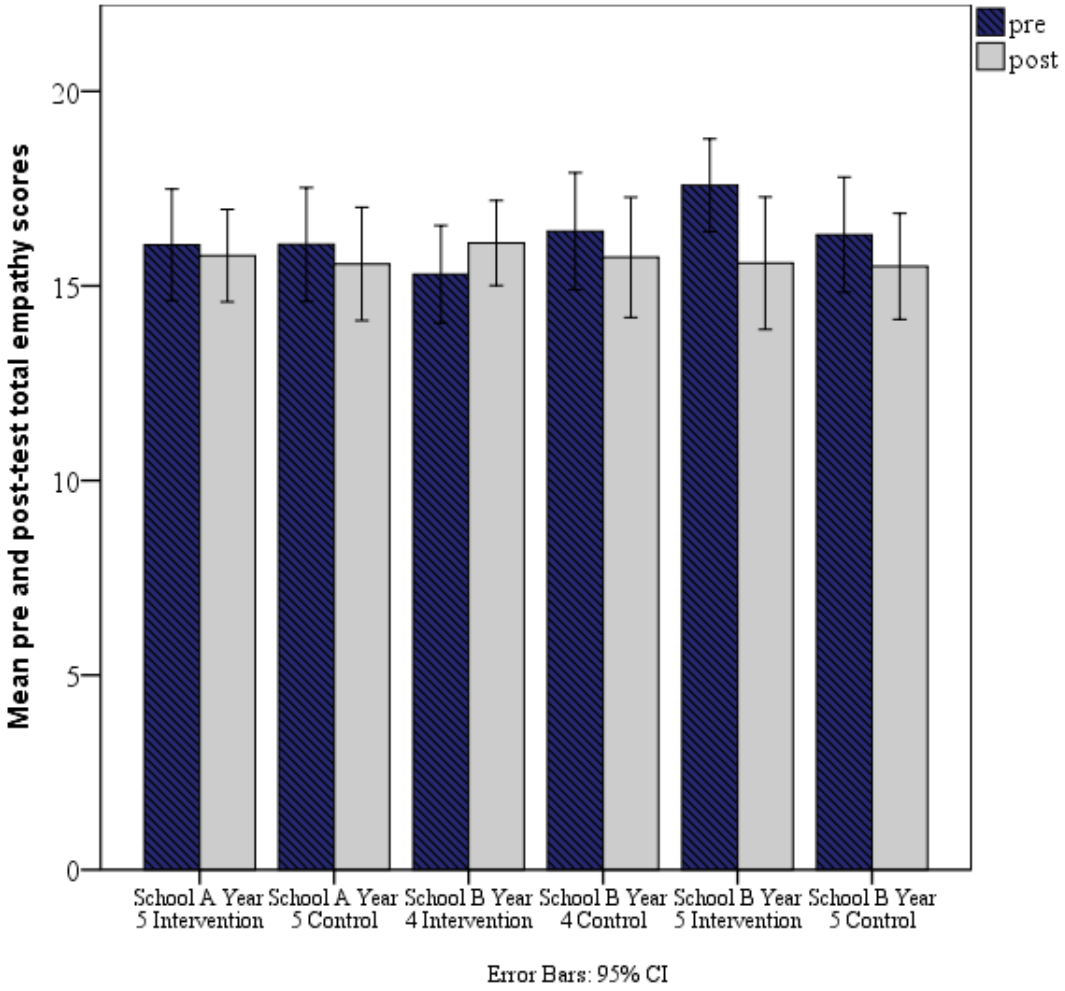


Figure 3.13 shows a similar pattern between the intervention and control group with a slight decrease in empathy scores in both groups with similar pre-test scores.

Table 3. 19 A table to show Mean, Standard deviation and gain scores pre-test and post-test on total empathy scores for each class.

Class	Intervention		N	Mean	SD	Gain score	SD
School A Year 5	Intervention	Pre	18	16.06	2.88	-0.28	2.35
		Post	18	15.78	2.39		
	Control	Pre	16	16.06	2.74	-0.50	2.73
		Post	16	15.56	2.73		
School B Year 4	Intervention	Pre	20	15.30	2.68	0.80	2.69
		Post	20	16.10	2.34		
	Control	Pre	15	16.40	2.72	-.067	1.91
		Post	15	15.74	2.79		
School B Year 5	Intervention	Pre	12	17.58	1.89	0.74	2.56
		Post	12	15.58	2.68		
	Control	Pre	16	16.31	2.77	-.081	1.97
		Post	16	15.50	2.56		

Figure 3. 14 A bar chart to show mean pre and post-test total empathy scores for intervention and control group in each class.



3.15.2 Empathy: Summary of Findings

There was a slight decrease in empathy scores in the School A Year 4 and School B Year 5, although this was small.. There was a slight increase in empathy scores in School Year 4 and a corresponding increase in the control group.

3.16 Bullying

3.16.1 Bullying index

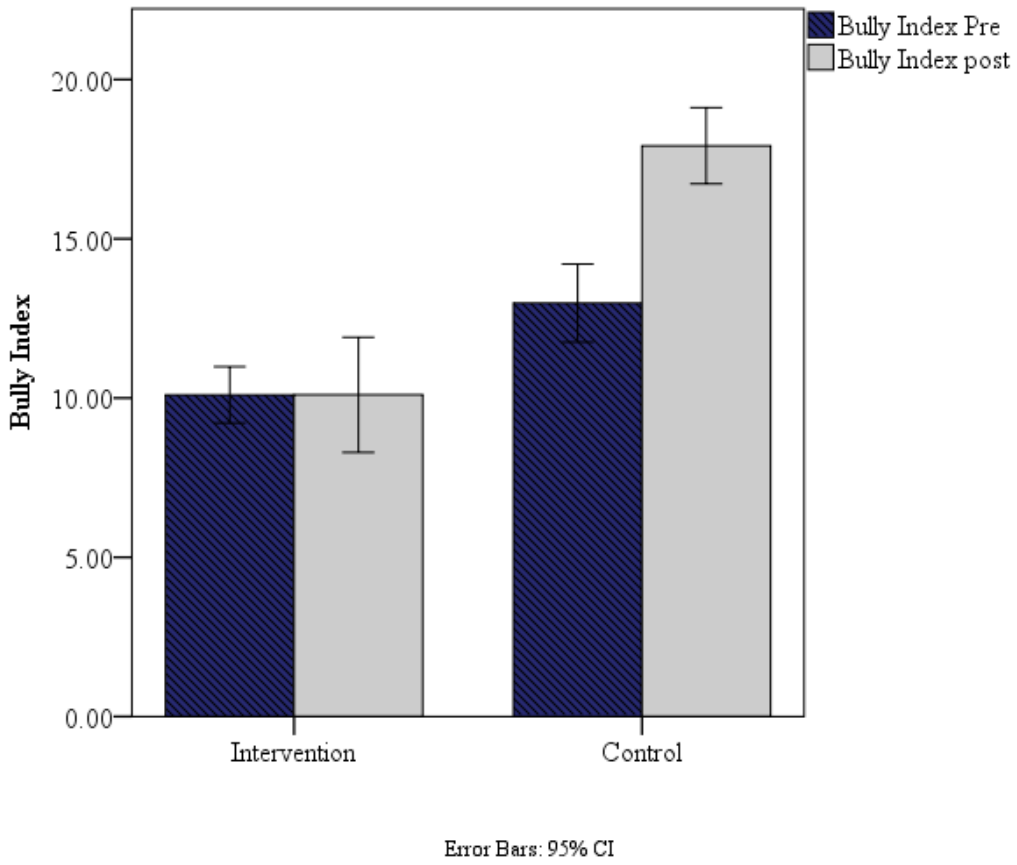
The Bullying Index represents the mean percentage of pupils that reported bullying during the current week on the My Life in Schools Checklist (Arora & Thompson, 1987). Section 2.7.3.2 outlines how the Bullying Index was calculated. The data used to calculate the Bullying Index can be found in Appendix T. Data for individual schools was combined for statistical analysis. Descriptive data for individual classes is presented, however statistical analysis was not conducted as Fredrickson and Cameron (2000) recommend that for groups smaller than 40 pupils the Index will be sufficiently reliable to allow valid comparisons. Limitations of this are addressed in section 4.2.5.2. The data used to calculate the Bullying Index and General Aggression Index can be found in Appendix U.

3.16.2 Descriptive Statistics

Table 3. 20 A table to show the mean and median for Bullying Index gain scores pre and post-test for the intervention and control group.

		N	Bullying Index (mean)	Bullying Index (median)	Gain Score (mean)	Gain Score (median)
Intervention	Pre	79	10.10	10.10	.004	-7.70
	Post	79	10.10	7.21		
Control	Pre	78	9.85	9.85	4.94	3.99
	Post	78	15.84	15.84		

Figure 3. 15 A bar chart to show the Bullying Index and mean gain scores at pre and post-test for the intervention and control group.



3.16.3 Statistical Analysis

As illustrated in Figure 3.15 there was an increase in the Bullying Index from pre to post-test in the control group with the Bullying Index remaining stable in the intervention group. There were no statistically significant differences at pre-test between the two groups (U=2862, Z = -.780, p = .435), however there was a statistically significant difference at post-test U=1352, Z = -6.157, p = .000) representing a medium effect (r = -.5. A Mann-Whitney U test also indicated that there was a significant difference in gain scores for intervention (Md = -7.70. n=79) and control pupils (Md =3.99, n=78) (U=2028, Z = -3.750, p =.000). This represented a medium effect (r = -.3). A Wilcoxon Signed Rank test indicated there were no statistically significant differences from pre to post-test in the intervention (Z = -.738, p = .460) with a very small effect (r = .08). However, there were statistically significant differences from pre to post-test in the control group (Z = -2.871, p = .004) with a medium effect size (r = .3). This was a statistically significant increase for the control group.

Table 3. 21 A table to show the Bullying Index and mean gain scores for pre and post-test for the intervention and control group in each class.

Class	Intervention		N	Bullying Index	Gain Score
School A Year 5	Intervention	Pre	24	14.91	-7.77
		Post	24	7.21	
	Control	Pre	24	21.05	-5.21
		Post	24	15.84	
School B Year 4	Intervention	Pre	29	10.61	-8.07
		Post	29	2.54	
	Control	Pre	28	8.97	3.99
		Post	27	12.96	
School B Year 5	Intervention	Pre	26	5.99	15.20
		Post	26	21.20	
	Control	Pre	26	9.85	15.32
		Post	26	25.18	

Figure 3. 16 A graph to show the Bullying Index and mean gain scores at pre and post-test for intervention and control groups in each class.

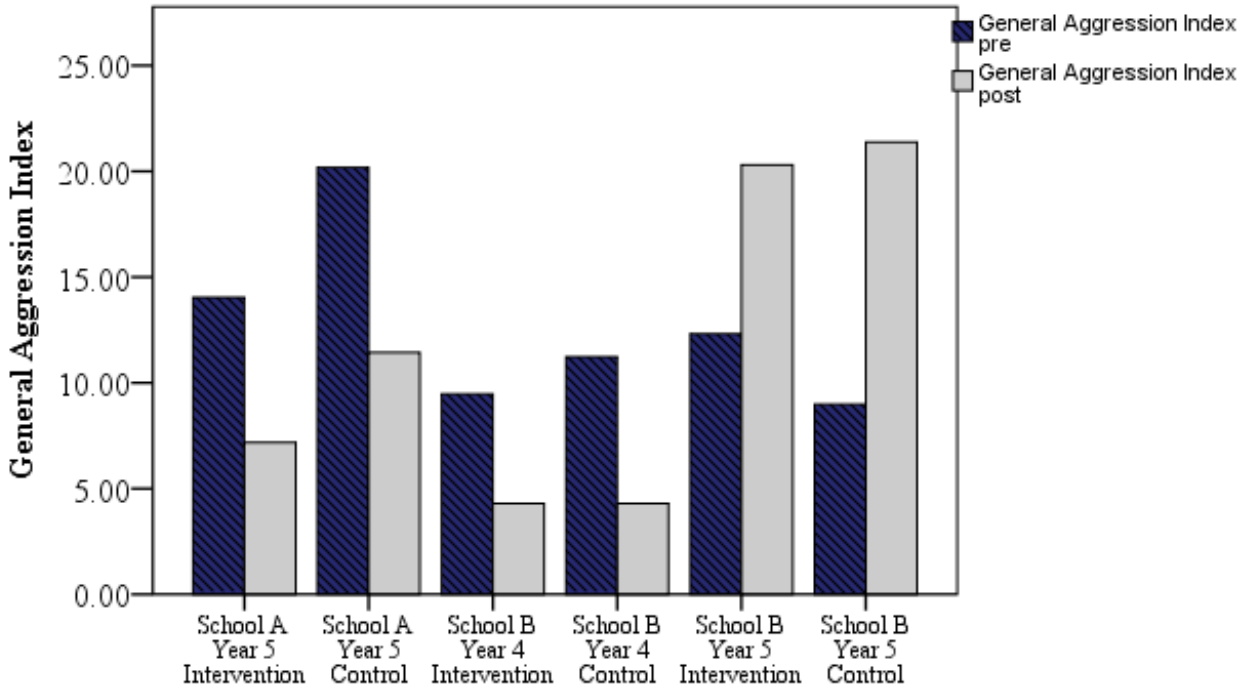


Table 3.21 and Figure 3.16 show a decrease in reported levels of bullying in the intervention group for School A Year 5, however, a similar trend was seen in the Control Group. The Bullying Index also decreased in the School B Year 4 intervention group with contrasting effects in the control group with an overall increase in reported bullying at post-test. In School B Year 5 there was a noticeably higher 'Bullying Index in both the intervention and control group at post-test.

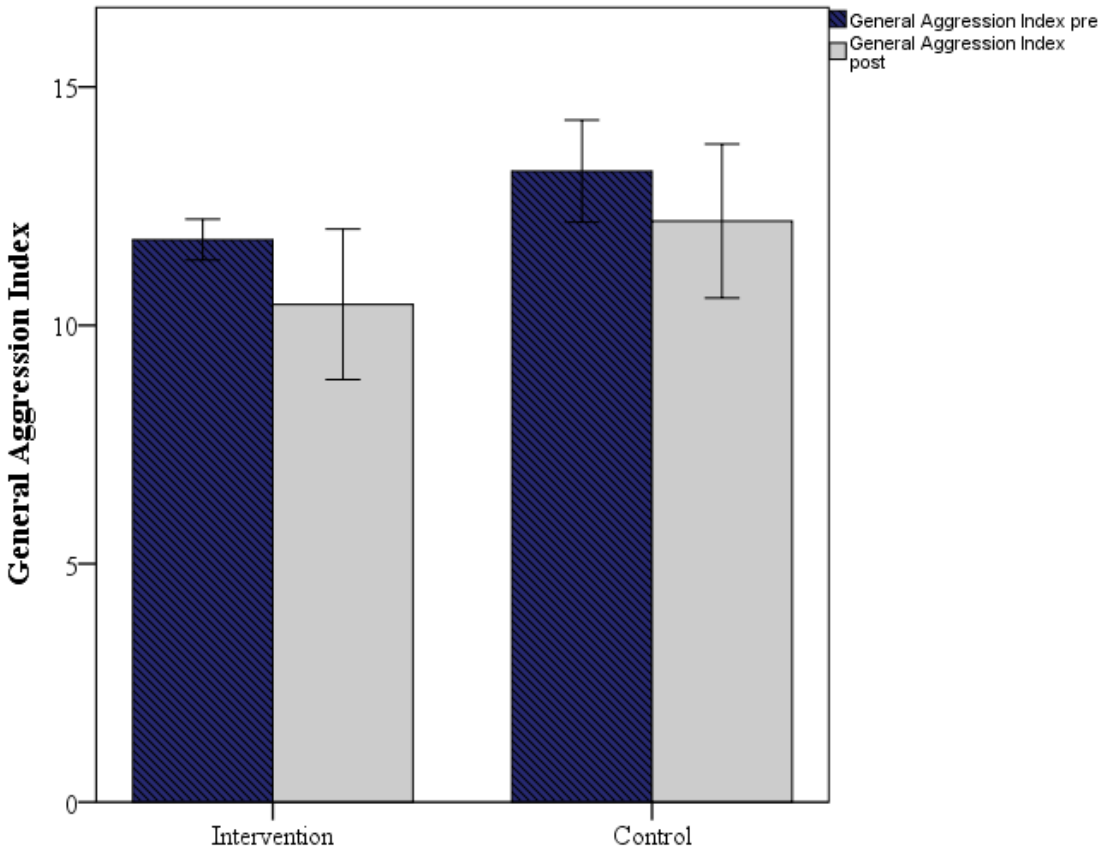
Table 3.21 also shows the amount of change in the Bullying Index in each class from pre to post-test. The negative changes indicate that the Bullying Index decreased between pre and post-testing, meaning that there was a reduction in the reported levels of bullying. A positive change would indicate that the Bullying Index increased from pre-test to post-test, meaning that there was an increase in reported levels of bullying. The gain scores indicate that there was a decrease in School A Year 5 and School B Year 4. The largest decrease was in school B Year 4 (-8.07) closely followed by School A Year 5 (-7.70). In school B Year 5 the increase was comparable in between the Cooperative Learning group (15.21) and (15.33) indicating that scores increased to a similar level across both the intervention and control groups at post-test.

3.16.4 General Aggression Index

Table 3. 22 A table to show mean and median for the General Aggression Index and gain scores pre and post-test for intervention and control group.

Intervention		General Aggression Index (mean)	General Aggression Index (median)	Gain Score (mean)	Gain Score (median)
Intervention	Pre	11.80	12.32	-1.36	-5.17
	Post	10.44	7.18		
Control	Pre	13.23	11.23	-1..05	-6.93
	Post	12.19	7.18		

Figure 3. 17 A graph to show the General Aggression Index and mean gain scores at pre and post-test for intervention and control groups.



As illustrated in Figure 3.17 there was an decrease in the General Aggression index from pre to post-test in the intervention and control group

There was no statistically significant differences between the two groups at pre-test (*U*=2708, *Z* = -1.33, *p* = .184). There was a statistically significance difference in the General Aggression Index between groups at post-test (*U*=2708, *Z* = -2.36, *p* = .018) representing a small effect (*r* = -.2) and statistically significant difference in gain scores between the two groups (*U*=2054, *Z* = -3.66, *p* = .000) representing a medium effect (*r* = -.3) but no significant difference between pre and post test scores in either the intervention (*Z* = .738, *p* =.460) or the control group (*Z* = -.821, *p* =.412).

3.16.5 Analysis by individual class

Table 3. 23 A table to show General Aggression Index and mean gain scores at pre-test and post-test for intervention and control group for each class.

Class	Intervention		General Aggression Index	Gain score
School A Year 5	Intervention	Pre	14.04	-6.86
		Post	7.18	
	Control	Pre	20.18	-8.75
		Post	11.43	
School B Year 4	Intervention	Pre	9.47	-5.17
		Post	4.30	
	Control	Pre	11.23	-1.07
		Post	4.30	
School B Year 5	Intervention	Pre	12.32	9.73
		Post	20.30	
	Control	Pre	8.98	12..43
		Post	21.38	

Figure 3. 18 A graph to show the General Aggression Index and mean gain scores at pre and post-test for intervention and control groups in each class.

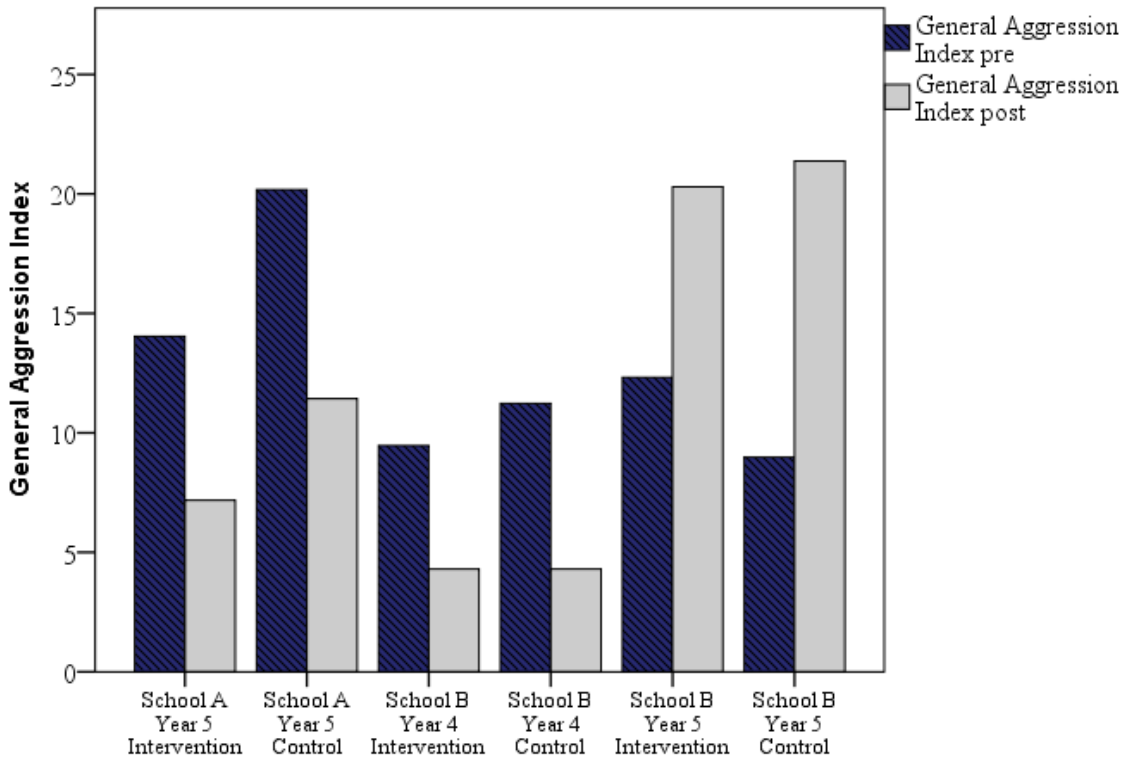


Table 3.23 and Figure 3.18 show a decrease in reported levels of aggression in the intervention group for School A Year 5, however, a similar trend was seen in the Control group. The General Aggression index also decreased in the School B Year 4 for the intervention and control group. In School B Year 5 there was a noticeably General Aggression index in both the intervention and control group at post-test.

3.17 Summary of Findings

Table 3. 24 A table to show a summary of findings, both overall and individual classes for each research question

Research Question	**Findings**	
	Overall	Individual Classes
What is the impact of CL on pupils' PA?	Play Same Gender (PDG): No statistically significant differences between intervention and control group gain scores (p = .266, r = .08).	PSG: No statistically significant differences in School A Year 5, School B Year 4 or School B Year 5 with small effect sizes (-.2, -.3 and -.2) respectively.
	Play Different Gender (PDG): No statistically significant differences between intervention and control group scores (p = .50, r = .2).	PDG: Statistical significance achieved in School A Year 5 (p = .034, d = -.6) . However, this was a statistically significant decrease in the intervention group.
	Work Same Gender (WSG): No statistically significant differences between intervention and control group gain scores (p = .621, r = -.04).	There were no statistically significant differences in School B Year 4 and School B Year 5 with small effect sizes (r = .09 and d = -.4).
		WSG: Statistically significant difference between gain scores in School B Year 4 (p=.013, r =.3) and between pre and post-test scores in the intervention group

	Work different Gender (WDG): No statistically significant differences between intervention and control group gain scores (p=.850, *d = .03).*	(*p*=.007, *r* .*5*) However, this was a statistically significant decrease. WDG: Statistically significant differences in School B Year 5 from pre to post-test the intervention (*p*=.003, *r* = -.*6*) and control group (*p*=.000, r = -7). However, this was a statistically significant decrease in both groups. There were no statistically significant differences in School A Year 5 or School B Year 4.
2. Does the impact of CL on PA differ depending on pupils' initial levels of PA?	PSG: Statistically significant difference in gain scores across the three levels of PA in the intervention group x^2 (2, n=79) = 16.89, p=.000 and control group x^2 (2, n=78) =8.70, p=.013. Follow tests indicated significant differences between low and high groups and medium and high groups in both the intervention and control group. Children with initially low levels of PA recorded an positive gain score compared to those with the high levels of PA who recorded a negative gain score PDG: Statistically significant difference in gain scores across the three levels of PA in the intervention group x^2 (2, n=79) = 15.55, p=.000 and Control group x^2 (2, n=79) = 16.40 p=.000. Children with initially low levels of PA recorded a gain score to those with a high level of PA who recorded a higher negative gain score in both groups.	

	WSG: Statistically significant difference in gain scores across the three levels of PA in the intervention group. (x^2 (2, n =79) = 10.982, p=.004). Children with initially high levels of PA recorded a high negative gain score compared to those with low or medium levels of PA. No statistically significant difference in gain scores across the three levels of PA in the control group (x^2 (2, n =79) = 4.120 , p =.127). WDG: Statistically significant difference in gain scores across the three levels of PA in the intervention group x^2 (2, n=79) = 15.224, p=.000. Children with initially low levels of PA recorded a positive change and children with initially high levels of PA recorded a high negative gain score. No statistically significant difference in gain scores across the three levels of PA in control group (x^2 (2, n=79) = 4.12, p = .445).	
3. What is the impact of CL on pupils' empathy	Gain scores in both the intervention (M = -0.26) and control group (M = -0.66) were both negative with little observable difference between the intervention and control group.	There was a slight increase in empathy scores in School B Year 4 from pre (M =15.30) to post (M =16.10) with a corresponding increase in the control group from pre (16.40) to post (M =15.74). There was a slight decrease in mean empathy scores in School A Year 4 from pre (M =16.06) to post (M =15.78) and School B Year 5, pre (M =17.58) to post (M =15.58). No analysis of statistical significance was conducted due to low-reliability coefficients.
4. What is the impact of CL on pupil's	Bullying Index: There was a statistically significant difference in the Bullying Index between groups at post-test (p =. 000, r =-.5) and a statistically	Decrease in the Bullying Index in the intervention group and control group for School A Year 5 and a decrease in School B Year 4 intervention group with an

experience of bullying in school?	significance difference in gain scores for intervention (p = .000, r = *.3)* with a statistically significant increase for the control group from pre to post (p =.004, r =.3) an increase in the Bulling Index. General Aggression Index: There was a statistically significance difference in the General Aggression Index between groups at post-test *(p = .018, r = -.2)* and a statistically significant difference in gain scores between the two groups (p = .000, r = -.3) but no significant difference between pre and post test scores in each group.	increase in the control group. In School B Year 5 there was a noticeably higher Bullying Index in both the intervention and control group at post-test. Decrease in General Aggression Index for School A Year 5 in both groups and School B Year 4 with an increase in both groups for School B Year 5.

3.18 Intervention Fidelity

3.18.1 Adherence

Four intervention checks were planned throughout the research and included in the research timetable see Appendix P. .However, due to staff absence not all of these checks were carried out in all intervention groups. Scores for each adherence check in each intervention class can be found in Table 3.24. A score of zero was given if the teacher was not present for the planned fidelity check.

3.18.2 Dosage

Actual intervention length for School A was less than the anticipated 10 weeks due to staff absence. Table 3.13 summarizes intervention dosage information collected from the teacher records. An example of completed teacher records can be found in Appendix S.

3.18.3 Intervention Fidelity in individual schools

3.18.3.1 School A

CL was used frequently, for example, up to 6 times in one week. There were, however, some interruptions through teacher absence. For example, in week 4 only one CL structure was used and there were no records provided for week 9 and 10. The use of the structures was varied, with CL being used in short bursts (e.g. 3 minutes) several times a lesson and over whole lessons (1 hour) or afternoons (2 hours). In terms of dosage, the total time spent on CL was 725 minutes which is equivalent to 12 hours 5 minutes over 8 weeks. This was the highest dosage out of all of the classes. When implemented intervention adherence was also high with an average of 90% from four fidelity checks, with key CL components incorporated and an observed emphasis on group processing with time to reflect on group skills at the end of the lesson. The structures which were used were varied and over the course of the intervention all five structures introduced were used multiple times.

3.18.3.2 School B

Year 4

Exposure to the intervention was varied throughout the intervention. The intervention ran for the specified 10 weeks, however in some weeks CL was only used for a total of 9 minutes for the week, and overall exposure was less than the other classes with 344 minutes (5 hours 44 minutes) spent on CL over 10 weeks. CL structures were varied, and all five structures were used during the intervention period with varying duration from short bursts (2minutes) to whole lessons (60 minutes). When implemented, intervention adherence averaged at 60% from three fidelity checks. One fidelity check was not completed due to the teacher not being available on the planned day. Fidelity checks highlighted that at times more emphasis was needed on the CL principles of Equal Participation and Simultaneous Interaction along with more reflection around small group skills. Pupils appeared engaged in the lessons and observations indicated that pupils started to know some of the names of the structures.

Year 5

The intervention ran for the specified 10 weeks; however, CL activities were typically used in short bursts, with the maximum length of sessions 30 minutes. In terms of dosage, CL was used for a total of 437 minutes (7 hours 17 minutes) over the intervention period. All five of the intervention structures were used during the intervention period. Intervention fidelity was low at 46%. This was partly due to staff absence which resulted in only two fidelity checks completed. However, fidelity checks indicated that groups were typically teacher selected and homogenous rather heterogeneous, although other CL elements were typically observed.

Table 3. 25 A table to show raw and average adherence score in percentages across four fidelity checks for each class.

			Check 2		Check 3		Check 4		Mean	Mean
	Raw	**%**	**Raw**	**%**	**Raw**	**%**	**Raw**	**%**	**raw score**	**%**
School A Year 5	43	93%	37	80%	45	93%	45	97%	43	90%
School B Year 4	37	80%	0	0	42	91%	32	70%	27.75	60%
School B Year 5	40	87%	0	0	0	0	44	96%	21	46%

Table 3. 26 A table to show a summary of the information relation to intervention dosage including duration, total number of CL sessions, duration of CL structures, time spent using in CL (in minutes) and the structures used.

	Dosage				
	Duration	**Total number of CL sessions**	**Duration of structures**	**Time spent using CL (in minutes)**	**Structures used**
School A Year 5	8 weeks	30	3 mins to 2 hours	725 (12 hours 5 minutes)	Think Pair Share Rally Coach Quiz Quiz Trade Numbered Heads Together Inside Outside Circle
School B Year 4	10 weeks	30	2 mins to 1 hour	344 (5 hours 44 minutes)	Think Pair Share Rally Robin Quiz Quiz Trade Numbered Heads Together Inside Outside Circle
School B Year 5	10 Weeks	29	5 mins to 30 minutes	437 (7 hours 17 minutes)	Think Pair Share Rally Robin Quiz Quiz Trade Numbered Head Together Inside Outside Circle

3.18.4 Summary of Intervention Fidelity

Overall, Intervention fidelity data highlights the diversity amongst the three groups in terms of how CL was introduced. The programme was intended to be delivered over 10 weeks; however, this was reduced in one class. Difficulties in undertaking fidelity checks in School B made it difficult to conclude whether there was any discrepancy between the concept and delivery of CL in the classroom. The data also highlights differences in the way CL structures were used with some classes incorporating the structures for whole lessons or afternoons and some preferring to use CL in short sharp bursts.

3.18.5 Anecdotal data

Anecdotal data was gathered and recorded throughout the research process from observations during fidelity checks, teacher comments on record sheets, and teacher comments and feedback during coaching sessions to aid the triangulation of the data. For the purposes of this summary Teachers will be given identifiers of A, B1 and B2. Teacher B2 was absent for two of the fidelity checks did not attend one coaching session and had to leave early from another therefore anecdotal data for this class is limited.

School A Year 5

Comments from Teacher A

During the first week of CL, Teacher A highlighted some concerns in regard to the introduction of social skills and managing the class as relationships between peers had been challenging historically noting that they *"struggled with kind words" and "struggled to work with others".* In particular, the teacher felt some pupils were finding this more challenging expressing that:

"A few of the children in the class have a diagnosis of ASC and are finding the positive comment element difficult".

Teacher comments indicated that even by week three some of the pupils were still finding working together a challenge with the teacher commenting that it is "*Tricky…they still had arguments"* and felt the structures were not that successful commenting that one structure (Rally Coach) *"didn't work" and "pupils need more work on coaching on how to guide a learner".*

By week four the teacher comments indicated that pupils still needed to be reminded about using elements such as *"positive praise"*, however despite this the teacher felt that *"children really enjoyed it"* (Quiz Quiz Trade) and noticed an impact on their learning commenting that *"they [pupils] retained more information".* This was echoed in week 5 with pupils showing *"good retention of knowledge and using knowledge the next day".*

During weeks 6 to 8 the teacher commented that children were beginning to say, *"they enjoyed working with new people"* and *"chose new people"* to work with and were

able to give *"lovely positive feedback"* to one another. During the coaching sessions, the teacher reflected on this further and commented that the use of *"heterogenous groups worked really well"* particularly with two boys with social communication difficulties. As the research progressed teacher comments and observations indicated that pupils were able to use and reflect on their social skills and were beginning to work cooperatively with less assistance from the teacher. Towards the end of the intervention, the teacher reported that:

> *"Children were suggesting using CL activities saying we could use Think Pair Share".*

During the last coaching session, the teacher reflected on the intervention. The teacher felt there was *"increased learning conversation between pupils"* and commented on the difference in the classroom expressing that:

> *"It's not World War 3 anymore, no one is throwing things, no one is going underneath the table".*

The teacher reflected on the positive impact of one pupil who has a diagnosis of ASC and reflected that *"previously he would bang his head on the table and have outbursts which meant the whole class had to leave the classroom"* The teacher commented that they *"haven't seen any of this behaviour since before October half term"*

As well as the behavioural difference in the classroom the class teacher commented on the social differences following using CL reporting*;*

> *"You can see the social differences between the two class [intervention and control], this class will accept to work with someone else now and will integrate with each other".*

The teacher also found there was a positive change in relation to bullying within the class commenting that there is:

> *"no bullying in the class anymore, whereas there was last year"* with pupils using the skills outside of the classroom *"trying to help solve the conflict on the playground".*

The teacher commented on the potential impact of CL on the pupils and expressed that:

"Pupils are happier and have increased self-esteem as they know that every day they are going to get a number of positive comments from their peers about their work"

In regard to implementation, the teacher commented that by the end of the intervention incorporating the structures became easier for example

"I can now just include the structures into most lessons and use them multiple times a day"

Pupil Comments

Pupil comments suggested that generally, they found CL positive when they were asked to comment on the activities at the end of the lesson, comments included *"Epic you learn new things" and "it was better than other literacy lessons because more people saw our work so there has been more input into it"*

Pupil comments during observations also suggested a more positive atmosphere commenting on the reduction of negative behaviours towards each other:

"It worked really well because we didn't argue and just got on with our work"

"We were more focussed and worked as a group and didn't argue".

"I thought it was good because it made learning fun and without arguing."

"It went really good because we all just got on with it, no one had a mardy and no one got in a grump".

"It went well because when we asked the questions we were calm"

Pupils comments also seemed to reflect some of the difficulties the teacher noticed particularly around pupils helping and coaching each other with one pupil commenting:

"At the beginning, I think we really struggled but when we got to the middle we really got the hang of it," the teacher asked why they think it was more difficult and the pupil expanded saying *"we weren't coaching each other to write an A grade sentence"*

Towards the end of the intervention period during the last fidelity check pupils positively reflected on working together:

> *"I think we worked well because we kept sharing ideas and took it in turns"*
>
> *"I think me and XXX worked well together giving information for each other's sentences"*

Pupils also spoke positively about working in heterogeneous groups and with people they would not normally work with for example:

> *"I thought it wouldn't go well because I thought [pupil] would bang the table and I would have to do all the work, but it actually worked well*".

Overall, pupils appeared engaged in CL and seemed to quickly learn the names, processes and requirements of each structure. Observations indicated a focus on group processing, social skills and interactions at the end of the lessons.

School B Year 4

Comments from Teacher B1

From the start of the intervention the teacher commented positively on the structures expressing *that "children loved it and it was quick/easy to implement*" and *"it [Rally Robin] resulted in more children joining in with high engagement"*

Over the course of the intervention, the teacher commented that pupils continued to *"like it"* and *"were all engaged"* and were able to use some of the social skills without prompting, for example, the teacher commented, *"They are doing more compliments more independently* `. The teacher spoke positively about the emphasis on social skills and commented that she would *"reflect on what good listening is"* and this seemed to have a positive impact for example *"this class then worked a lot better than the other class"* and felt by using the structures the *"class seem to get a lot more done and a lot less chit chat, more productive"*

Despite these positive comments, the teacher did note that *"some worked better than others"* and during the coaching session expressed that it was *"more difficult with peers with the higher ability"* as well as commenting on potential difficulties

implementing the Quiz Quiz Trade structures due to *"not having time to prepare the quiz cards"*

The teacher felt that by using CL the "pupils are becoming more accepting of each other" and the structures enable pupils to work together more effectively commenting that *"If they disagree it's not the end of the world, they can express an opinion"* and gave an anecdotal description of two pupils who had previously had some difficulties working together and were grouped together by an outside agency. The teacher noticed them using some of the skills taught in the lesson one pupil actively tried to include the one another for example one pupil said, *"hold on [pupil] has a different idea, lets listen to her idea"* and the pupils were more accepting of different opinions.

School B Year 5

Teacher B2 comments

The teacher commented that even at the start of the intervention the pupils seemed to like the CL structures and *"were excited to start the task and liked listening"* and felt that the use of social skills improved over time for example *"was interesting to see who was listening and their active listening skills, which were better than last time".*

During coaching sessions, the teacher commented that she was finding it a challenge to implement some of the structures due to the class being in sets for some lessons and her lessons often clashing with P.E and planning time. However, despite some challenges in implementing the teacher commented that she felt the pupils *"are nicer, friendly and not just saying their own opinion and are thinking about others spontaneously".*

3.19 Conclusion

Overall, the teacher comments highlight that there were some challenges at the beginning of the intervention with children struggling with some of the social skills needed to work cooperatively with pupils needing to be reminded of these skills, particularly during the first four weeks. There were some additional implementation challenges highlighted, including difficulties with peers with higher abilities and having time to prepare resources. However, as the intervention progressed comments

indicated that children started to work with cooperatively with new people with noticeable differences in behaviours and social skills between the intervention and control group with less bullying reported and an increased ability to solve conflict and reflect on social skills. Pupils were also reported to be more accepting of other opinions and were considering the opinions and feeling of others. Pupil comments indicate that they could notice positive differences in their interactions and were able to reflect positively on working with new people. Comments also highlighted that some teachers felt they were able to easily implement the structures with relative ease by the end of the intervention.

4 Discussion

The aim of this research was to evaluate the impact of CL on PA, empathy and pupils experience of bullying in school. This chapter outlines the findings of the research, considering each research question in turn. The findings are critically discussed with considerations to relevant literature and theory. A review of the methodology and measures used is presented followed by a discussion of the implications for EP practice and suggestions for future research. Finally, the unique contribution of the research is discussed.

4.1 Research findings

Research questions one and two are related to pupils' PA which is measured by the SIS (Fredrickson & Furnham, 1998). Research question three is related to perceptions of empathy as measured by items from the emotional literacy scale (Faupel, 2003) and question four relates to experiences of bullying in school, as measured by the My Life in Schools Checklist (Arora & Thompson, 1987). The findings from each research question will now be considered in turn, with links made to relevant literature.

4.1.1 Research Question One: What is the impact of CL on pupils' PA?

When data from all schools are combined, CL did not produce positive gains in any of the four contexts (PSG, PDG, WSG, and WDG). However, there were significant pre-test differences in School B and with this removed there was a statistically significant difference between gain scores in the intervention and control group in the PSG context, representing a small effect (r = - .2).

At an individual class level there were no statistically significant findings in the PSG context, however in the PDG context statistically significant differences were found in School A Year 5, although this was a statistically significant decrease in the intervention group and represented a medium effect. In the WSG context, there were statistically significant differences in gain scores in School B Year 5, with a medium effect, however again this was a significant decrease for the intervention group. Similarly, in the WDG context, statistically significant decreases with a medium effect were found in School B Year 5 for the intervention group.

4.1.2 Research Question Two: Does the impact of CL on PA differ depending on pupils' initial levels of PA?

In considering a potential differential impact, the findings suggest that the impact of CL may vary according to pupil's initial levels of PA. Findings indicated gains for those with initial low levels of PA in the intervention group compared to the control group for the PDG and WDG context. Findings also indicated that in the intervention group for the PSG, WSG and WDG context the PA for those with initial higher levels of PA reduced. These findings could be interpreted to suggest that CL has the most impact on those with initially low and high levels of PA.

4.1.3 Links to the literature: PA

The possible decreases in PA could be compounded by inconsistences in the implementation of the intervention, resulting in a negative effect. Battistich, Solomon, & Delucchi (1993) indicate that the effects of CL depend upon the quality of the interaction, with findings suggesting that frequent low-quality interactions are associated with negative outcomes. Battistich, Solomon, & Delucchi (1993) found some students work in groups relatively often but do not work together well, arguing that this combination does not produce positive outcomes. The data collected in regard to implementation fidelity suggests that some classes used CL frequently but for short periods of time and observations indicated in some instances' interactions were of a low quality. These findings could be interpreted to indicate that simply just implementing CL in the class, with little regard to the quality of the interactions could have a negative impact. Although, intervention fidelity data in School A indicated a good level of adherence and positive interactions were seen through observations and reflected in teacher and pupil feedback, which does not seem to fit in with the above possible explanation. A further reason for the lack of effect could be related to the measure used as it may not have been sensitive enough to pick up subtle changes within the classroom ecology.

A further barrier and explanation for the negative effect found could be related to the school ethos and system. The rationale for the research was based theoretically on social interdependence theory and the principle of positive interdependence. Positive interdependence (when the actions of individuals promoted the achievement of joint goals) is thought to lead to promotive interaction which was hypothesised to make

group members feel more accepted, more willing to help others in return (Gillies, 2016) and re-evaluate the social desirability of others (Bierman, 2004). However, it could be argued that social interdependence is not enough, as it may be possible for pupils to work together without necessarily developing supportive relationships (Hymel et al., 2015).

This raises the question whether simply using CL in selected classes within a school system is enough to promote positive outcomes. CL is likely to be influenced by the school ethos which could create a barrier if this is fundamentally incompatible to CL. There is, therefore, the need to create the right climate before trying to implement CL in a classroom or school. Future research should therefore have greater consideration of the wider school ethos and culture before implementing CL.

It could also be argued that looking at overall PA does not give the whole picture. Research into peer ecology raises the question as to whether an overall increase in PA across the classroom is something that should be seen as a positive outcome. The status of peers within a classroom can be structured differently and it is proposed that some characteristics, such as social structure, only exist at the classroom level and cannot be understood by student level outcomes (Hendrickx, Mainhard, Boor-Klip, Cillessen, & Brekelmans, 2015).

Classrooms can be egalitarian (where each student is equal) and hierarchal (where a small number of students have a relatively high peer status and a small number with relatively low peer status) (Hendrickx, Mainhard, Boor-Klip, Cillessen, & Brekelmans, 2015). By splitting pupils into low, medium and high categories, although arbitrary and accompanied by its own limitations, it begins to explore whether CL impacts upon the hierarchy of the classroom. The findings of the research indicate that in some contexts PA scores increased in those with initial low levels of PA and decreased in those with high levels of PA. This could be seen as reducing the status hierarchy within the classroom, which in turn could create more equal status amongst pupils. Although viewing a decrease in PA for those who were initially accepted by their peers could be controversial, research into social dominance and status hierarchy could provide an argument for this. Research has shown that in a classroom with a pronounced status hierarchy there were more negative behaviours, related to bullying

and victimisation, than in a classroom where social status was evenly distributed (Schäfer, Korn, Brodbeck, Wolke, & Schulz, 2005).

Traditionally, status hierarchies have been seen to decrease intra group conflict by increasing the predictability and stability of social relationships (Pellegrini & Long, 2002), research is beginning to challenge this perspective in favour of supporting a "balance-of-power" view. From this perspective it could be argued that the power differential which is inherent in hierarchical contexts is detrimental to peer relationships, with status hierarchy in adolescence found to be associated with higher levels of bullying (Garandeau, Lee, & Salmivalli, 2013). Individuals of average status are also less likely than low and high-status individuals to engage in aggressive or defiant behaviours (Closson, 2009). This study therefore highlights further questions to be explored in future research in relation to the relationship between classroom ecologies and PA and the mechanisms by which CL impacts PA within the classroom.

4.1.4 Research Question Three: what the impact of CL on pupils' empathy is

With data from all schools combined, data was relatively stable from pre to post-test between the intervention and the control group. In terms of individual classes there was a slight increase in scores from pre to post-test in School B Year 4 and a slight decrease in the other two classes. However, these findings are descriptive and should be interpreted with the acknowledgement of the low reliability of the measure.

4.1.5 Links to the literature: Empathy

The role of empathy in intergroup relations is complex as empathy can take a variety of forms and therefore be encouraged in several ways (Stephen & Finlay, 1999). In the CL intervention empathy was not explicitly taught, and despite suggestions that the induction of empathy should not be left to chance (Stephen & Finlay, 1999), it was hypothesised that through working interdependently with those from other groups children will learn to view the world from the perspective of others (Bridgeman, 1981). Piaget (1965) asserts that dialectical peer interaction, which is encouraged in CL, promotes the ability to take the perspective of others, with perspective taking linked to empathy (Batson, Early, & Salvarani, 1997) with the terms used interchangeably within the literature.

The lack of change in empathy scores could also be related to the validity of the measure. Stephen & Finlay (1999) used Bridgeman's (1981) study as an example of one of the few that has used empathy as a mediating or outcome variable. However, Bridgeman's research looks at role taking, measured by Chandler's (1973) cartoon series, and Moral Reasoning, measured by Kohlberg's (1963) moral dilemmas. Therefore, it could be argued that this was not measuring empathy in the same way described by other researchers with Cuff, Brown, Taylor and Howat (2016) identifying 43 distinct definitions and conceptual summaries describing empathy. In recent research Stietz, Jauk, & Krach (2019) argue that empathy should not be used as an umbrella term which encompasses all routes to understanding others; they argue for its disassociation from other social cognition such as perspective taking. It is therefore possible that the measure in this study was not tapping the right construct for an observable difference to be found. Further discussion of the empathy measure is presented in in Section 4.2.5.3.

Moreover, it could be that empathy is not a prerequisite skill for prosocial behaviour and the development of empathy could be related to increases in PA. Social judgement theory, discussed in section 1.8.4.1, suggests that contact between diverse individuals increases liking or hostility towards each other and it could be argued, therefore, if promotive interaction has not taken place and liking amongst peers has not increased, as PA findings indicate, then it is unlikely that perspective taking and empathy will increase.

4.1.6 Research question four: What is the impact of CL on pupil's experience of bullying in school.

This question was addressed using the My Life in Schools Checklist (Arora and Thompson, 1987). Despite a small decrease in the median the Bullying Index remained stable within the intervention group with an increase in the control group. There were, however, decreases in the General Aggression Index in the intervention group, which could be in line with Arora & Thompson (1987) who state that a reduction in aggression is likely to be identified before a decrease in bullying. There was, however, a similar decrease in the control group.

4.1.7 Links to the Literature: Pupils experiences of bullying in school

CL was hypothesised to influence bullying through the promotion of interdependence. Social interdependence theory would argue that since in CL contexts individuals are more likely to be equal (Choi, Johnson, & Johnson, 2011) the reduced hierarchy within the classroom means pupils are less likely to be bullies or victims. There was a significant difference between gain scores in the intervention and control group, and although there were not significant increases in the intervention group it could be interpreted that bullying within the class at least stabilised within the intervention group compared to the control group.

Anecdotal data from School B suggests that towards the end of the intervention the teacher felt bullying had decreased compared to the previous year. It is possible that intervention needs to be longer in duration in order to see an observable effect in self-reported measures. Although, the measure's sensitivity to subtle changes within the classroom peer ecology could also be questioned. A review of the measure is presented in section 4.2.5.2.

Although data should be interpreted with caution, this data indicates further research is needed to explore the links between CL and bullying. Research indicates that bullying should be understood within a social-ecological framework (Swearer Napolitano, Espelage, Vaillancourt, & Hymel, 2010). Rates in bullying vary as a function of the school climate (Wang, Berry, & Swearer, 2013) with students more likely to participate in bullying when the school climate is perceived as negative (Gendron, Williams & Nancy, & Guerra, 2011) and norms promote bullying (Unnever & Cornell, 2003). Bullying, therefore, is not only influenced by the social dynamics of the immediate and wider peer group but also the school system and the ethics of the wider community (Fredrickson & Cameron, 1999). Consequently, it is argued that bullying is best tackled through intervention at a group and organisational level (Fredrickson & Cameron, 1999). This view is synonymous with more recent research which suggests that anti-bullying initiatives need to shift group norms and dynamics in order to create a community in which bullying is less likely to happen (Hymel et al., 2015). Therefore, it could be argued that positive social interdependence alone cannot be relied on to decrease bullying (Hymel et al., 2015) and an intervention must also focus on the wider ecology and school climate.

4.1.8 Procedural Aspects

The Kagan approach (Kagan & Kagan, 2015) was selected as the model of CL. It was chosen over other models as it relies on simple structures which are not tied to a curriculum, requires no special materials or preparation and no change in lesson design or content (Kagan & Kagan, 2015). However, despite these features and the structures available within it, there was some evidence of teachers placing less emphasis on come core elements, potentially reducing impact.

A key component of the approach, social skills is emphasised within the Kagan model and wider literature (e.g. Gillies, 2016; Goodwin, 1999). Social skills were emphasised within the intervention training, with steps provided to teach students interpersonal and group skills with students engaging in a distinct period of social skills training prior to CL. However, no fidelity checks were undertaken during this period therefore the focus on social skills could have varied between groups. Teacher records for one class also indicate that one teacher did also try and introduce CL during this period. During the intervention period a fidelity check showed that a focus on group processing was usually present, however this was not consistent across all classes. Future training needs to be adapted to give greater emphasis on this element and fidelity checks should be undertaken during the social skills period to support further emphasis on the use of social skills.

4.2 Methodological Review

A post-positivist quasi-experimental methodology (section 2.2) allowed for causal relationships to be explored in a real-world setting. However, conducting research within a real-world setting was accompanied with a number of challenges.

4.2.1 Internal Validity

4.2.1.1 Selection

Quasi-experiments often include non-equivalent groups (Shadish, Cook, & Campbell, 2002) and here there were some existing differences at pre-test in School B Year 5 compared to other classes. A further potential threat to internal validity is the contextual difference between schools meaning that an observed difference may be related to initial or ongoing (‘history threat’) differences between the groups. Control

groups were in the same school as the intervention classes, creating a risk of diffusion of treatment from the intervention to the comparison group. In addition, intervention pupils in School B worked with those in the control group, as classes were mixed for English and Maths lessons based on ability, potentially diluting the study.

4.2.1.2 Sample size

Power calculations (section 3.4) indicated that there may not have been sufficient power to detect a statistical effect in analysis at the level of the individual school. Replication of this study would be valuable with a larger sample.

4.2.2 Regression to the mean

Regression to the mean is a statistical phenomenon that affects all pre-experimental designs. Scores towards the ends of the distribution will, on average, be more likely to have higher error term than those nearer the mean. At re-test the results towards the end of the distribution will tend to move closer to the mean than those in the middle range. Marsden & Torgerson (2012) assert that if participants' gain scores indicate a negative change it could mean that regression has occurred. This risk is pertinent when investigating the *differential* impact of CL upon prior PA. Increases for the low PA group in comparison to the high PA group therefore need to be interpreted with some caution. Nevertheless, this pattern was not consistently observed in the control group and could indicate that effects found were from the CL intervention.

4.2.3 Intervention fidelity

The duration of the intervention was 10 weeks, selected for pragmatic reasons: term length and researcher capacity. However, this is considered to be relatively short within the literature (Murphy et al., 2005). As Cohen, Manion, & Morrison (2011) assert, a lack of positive outcomes in quantitative measures may indicate that the intervention was of too short duration for an impact to be measured rather than indicating that an intervention has not been successful.

The absence of inter-rater reliability for the fidelity checks remains a limitation along with the absence of some fidelity checks for School B. Intervention fidelity for this school, therefore, cannot be assumed which could have impacted upon the findings of the study.

Additionally, no observations were carried out prior to the intervention period commencing. This is in contrast to advice as it is suggested that before beginning intervention delivery, researchers should assess providers' acquisition of skills by written test, direct observation, or a combination (Bellg et al., 2004).

In order to ensure that there was no contamination across the two groups two fidelity checks were carried out in the control groups at the beginning and end of the research to identify whether any CL elements were being used and to establish whether there was any likely diffusion of treatment. Scored for fidelity checks Scores for fidelity checks carried out in the control group can be found in Table 2.5 in section 2.6.8.

4.2.4 External validity

The impact of CL was not the same across all classes therefore caution should be exercised in attributing the observed effects solely to the intervention itself as it is possible that other contextual factors influenced the findings. For example, the Bullying Index decreased in both the intervention and control group in School A and School B year 4 but there were increases in both groups in School B Year 5, which could indicate that there were other environmental influences occurring.

4.2.4.1 Reliability

Fluctuations in pupils' responses from pre to post-test may be due to influences other than the intervention. In the PA measure, for example, it is possible that pupils may have responded positively or optimistically at pre-test but after having an opportunity to work with their classmates, at post-test responded more '*realistically*'. The CL structures encourage heterogenous groups, with teachers trying to group peers who would not typically choose to work together meaning that pupils were often paired with less familiar peers. This experience may have afforded pupils the opportunity to make more realistic judgements about whether they would like to work or play with others.

The empathy measure also revealed some difficulties in relation to reliability. Items were read out loud to ensure comprehension, but some children seemed to find the reversed and negatively worded questions difficult to interpret. Whether pupils accurately and reliably selected the option intended was therefore difficult to ascertain.

In future small group administration of all measures, in particular the empathy measure, would be useful in order to enhance the reliability of responses.

4.2.5 Review of measures

4.2.5.1 PA

The SIS (Fredrickson & Furnham, 1998) is a sociometric measure, which could be influenced by events immediately preceding the completion of the measure and may therefore not give a representation of typical classroom dynamics. Vaillancourt, Hymel, & Mcdougall (2001) also assert that sociometric liking/disliking is not necessarily synonymous with perceived status or popularity: being rejected is not the same as being viewed by peers as unpopular or low in status (Vaillancourt et al., 2001). Therefore, future research could use direct observations of behaviour to explore students' interaction patterns.

4.2.5.2 My Life in Schools Checklist

The My Life in Schools Checklist (Arora & Thompson, 1987) asks questions about different experiences in school and avoids asking pupils directly whether they are being bullied. This is a strength of the measure as the word bullying could raise ethical challenges or be perceived in different ways. As a further strength, pupils are asked to report events from within the last week. This means that information reflects current experience and current peer ecology.

The measure does, however, have limitations in that the six items used to calculate the Bullying Index describe verbal and physical acts of bullying; but do not reflect more indirect forms of bullying. This seems crucial to understand given that girls, for example, are more likely to experience this type of bullying (Carbone-Lopez, Esbensen, & Brick, 2010). Additionally, the measure has not yet been updated to include new forms of bullying such as cyberbullying. The checklist is limited by threats in self-report measurement (Van De Mortel, 2008) and low reliability when calculating the Bullying Index and General Aggression Index for groups of less than 40 (Fredrickson & Cameron,1999).

4.2.5.3 Empathy

Empathy is described as an elusive construct, difficult to define and measure (Warden & Mackinnon, 2003). This is compounded by difficulties in accessing measures used in previous research, as they are not widely available. As outlined in section 2.7.2.1 the empathy measure in this study was adapted from the Emotional Literacy Scales pupil checklist (Faupel, 2003) and the re-calculated alpha coefficient (.4) indicated low reliability. Further research needs to consider the development of a reliable and accessible empathy measure.

4.2.6 Review of research design aspects

The research adopted a post-positivist standpoint in an attempt to isolate, control and manipulate variables in order to understand causal relationships. However, this allowed for limited exploration of the mechanisms underpinning any effects. A constructivist qualitative approach may be of interest in the future to allow for a flexible exploration into the experiences of pupils taking part in CL.

Data was analysed using gain scores. Limitations associated with this approach are acknowledged in section 3.2.1. Additionally, in exploring the differential impact of the intervention upon PA the interquartile range was calculated for the whole group rather than individual class groups. This grouping method increases the threat of regression to the mean and ceiling effects.

Furthermore, the interaction between the measures was not explored as this was beyond the scope of the research question but could be a topic for future exploration. Recent research has shown that marginalised students reported significantly lower levels of bullying, victimisation and perceived stress compared to those in control schools (Van Ryzin & Roseth, 2018). Within this research initial PA levels were not correlated with experiences of bullying, serving as a potential limitation and an area in need of further attention.

Similarly, although intervention fidelity was observed, and data was collected correlations between intervention fidelity and outcomes were not explored. This is something that should be explored in future research.

4.3 Implications of the findings

4.3.1 Implications for practice

The study highlights implications for the role of the EP at a systemic level through providing initial training and ongoing support to teachers to deliver educational innovations. Through working in this way EPs have the opportunities to promote positive outcomes for a greater number of children and young people, not just those with special educational needs. Furthermore, there is a need for EPs to consider the wider context within an organisation and how this may impact the successful implementation of initiatives and interventions. Within the study the researcher had limited involvement with senior leaders, with intervention at a class level, which may have contributed to the findings given that leadership involvement is regarded as important for initiatives to be successful (Bosworth, Garcia, Judkins, & Saliba, 2018). With some EPs able to move towards more systemic ways of working they are in a good position to work collaboratively with schools ensuring staff readiness before implementation, promoting effective practice and developing implementation standards within Local Authorities (Moir, 2018).

The negative effect found in some PA contexts indicate that evidence-based practice requires attention to context and intervention features: notions captured in implementation science (Gulliford, 2015). Fixsen, Blase, Horner, & Sugai (2009) note 'choosing an evidence-based practice is one thing, implementation of that practice is another thing altogether' (p.5). Implementation is the critical link between research and practice (Cook & Odom, 2013) and educational innovations such as CL may be no more successful than other teaching methods unless time is devoted to implementation (Topping et al., 2011). Despite the increased emphasis on implementation science within education (Lyon et al., 2018; Moir, 2018) the changing educational landscape puts pressures on schools and services to make positive changes quickly in the most effective way with factors involved in implementation often overlooked. Implementation is argued to be the next and most critical stage of evidence-based practice (Cook & Odom, 2013). EPs, who are the link between the worlds of academic psychology and education (Elliott, 2002) and act as scientist practitioners, need to play a role in bridging the gap (Carnine, 1997) through promoting implementation science alongside evidence-based practice.

A further implication for practice in terms of CL is how models such as Kagan can be incorporated into classrooms. The use of Kagan structures in this study seemed to provide an accessible introduction to CL. Although teachers' views of the model of CL and structures were not explicitly explored in this study the anecdotal data gathered indicates that the use of Kagan structures can provide an accessible introduction to CL, which teachers feel is sustainable in practice. Some deviations from the model still occurred, however, further highlighting the need for continued support to ensure CL is implemented with high fidelity

4.3.2 Implications for future research

The research has four main implications in terms of theory, measurement, CL and implementation.

First, the research has explored the impact of CL on pupils' PA, empathy and pupils' experience of bullying in school. The results of this study found an effect in some contexts which was not in the direction expected, therefore based on these findings caution needs to be taken when considering recommending and implementing CL in UK classrooms, with further consideration given to its applicability given its USA origins. Given the current findings of the study CL would not be recommended as an approach to increase peer acceptance, empathy and pupil's experiences of bullying. Further research is needed within this area, which addresses some of the methodological limitations of this research.

Although the research found an effect in some contexts which was not in the direction expected, the study went some way in exploring for who CL most effectively enhances PA. Findings in relation to a potential differential impact pose important questions about the nature of PA and the ecology of the classroom, which has important implications for theory and practice. It could be hypothesised that a linear viewing of PA, with increases across the whole class, may not be in line with research into peer ecology. Egalitarian classrooms, as opposed to hierarchical, promote positive outcomes (Rodkin & Gest, 2010) with less victimisation and bullying (Schäfer et al., 2004). Increases across the whole class are therefore likely to keep the status hierarchy of the classroom intact, which could negatively impact the peer ecology. Therefore, there needs to be further consideration of the nature of PA and the ecological structure of the classroom. Further research needs to consider why and how CL might influence

PA levels. This also has implications for research design and analysis as individual response to interventions may be masked by the dependence on measures of central tendency, reducing insight into the effect of an intervention. There is also a need to explore the relationships between the three constructs of PA, bullying and empathy both theoretically and in research.

Second, the study highlights the need for a reliable and valid measure of empathy in children and young people to be developed which is easily accessible and can be used in educational research. Additionally, the measurement in the current study did not provide information into how relationships between peers evolved over the term. This could be addressed through more detailed socio-metric methods such as peer nomination sociograms supplemented by coded interactions between peers.

There are a number of avenues which would benefit from further consideration, including further exploration of how CL impacts on peer acceptance and other aspects of the peer ecology, how CL could be incorporated into a secondary schools curriculum for year 7 pupils and the impact of whole school CL. These areas for future research are summarised in Table 4.1.

Finally, the fact that CL is a complex pedagogical approach is not in doubt: close attention is required to the integrity and fidelity to support inferences from research. Future research needs to consider research from implementation science (Moir, 2018). There is a greater need to explore the mechanisms which make CL effective and the factors implicated in the implementation process which can be used consistently across CL research. This could also explore the key elements in CL by contrasting different models, training and support structures, grouping strategies and the use of social skills. There is also a need to develop pragmatic and reliable procedures to support intervention fidelity in CL research with clearer guidelines as to the optimum length of intervention, dosage and use of structures.

Table 4. 1 Indications for future research

Potential future research questions	Rationale	Possible Designs
How does CL influence peer acceptance?	A linear viewing of PA, with increases across the whole class, may not be in line with research into peer ecology. Further exploration as to how CL influences peer acceptance and the mechanisms involved is needed.	Quasi-experimental Coded Observations of interactions Use of sociograms
How does CL impact other aspects of the peer ecology?	Drawing on theories of social hierarchy, further exploration of the impact of CL on aspects of the peer ecology such as classroom norms, social behaviour and richness of interpersonal ties using sociometric measures to not only explore the degree to which classmates liked each other but also the degree to which classmates viewed peers as popular.	Quasi-experimental Coded Observations of interactions
What is the relationships between the three constructs of PA, bullying and empathy?	Further exploration investigating the relationship between the three constructs, as bullying has been associated with low empathy (Endresen & Olweus, 2001) and low PA (de Bruyn et al., 2010) with PA thought to act as a protective factor against bullying and victimisation (Pellegrini & Bartini, 2000).	Quasi-experimental Correlational
What are the barriers to and facilitators of the implementation of CL? **To what extent are CL outcomes influenced by Intervention Fidelity?**	Implementation Science research aims to understand what contributes to the successful implementation of CL (Forman et al., 2013) as even with solid foundations in theory and empirical	Quasi-experimental Coded video observations Correlational

What is the optimum intervention length and optimums frequency and duration of structures for CL to have positive outcomes?	research, interventions can still yield poor results due to poor implementation (Moir, 2018). Taking a post-positivist approach would illuminate the key factors in implementing CL, aiming for control of intervention features.	
How can CL be incorporated into a secondary school curriculum for year 7 pupils?	Secondary schools are very different organisational environments (Topping et al, 2011) and positive findings from primary schools cannot be assumed to translate into secondary schools. The transition from primary to secondary school is critical (Hanewald, 2013) and peers can play a pivotal role in this transition. It would, therefore, be useful to see how CL can be implemented within a secondary school setting.	Quasi-experimental
What is the impact of whole school CL on PA, empathy and pupil's experiences of bullying?	Ignoring the whole system of which the school is part of is a threat to an intervention (Moir, 2018). There is a need for further exploration of the impact of CL when it is incorporated into the macrosystem, for example as a whole school approach. This is supported by suggestions that positive interdependence may not be enough to facilitate supportive relationships (Hymel et al., 2015) and the suggestion that bullying is best tackled though intervention at a group and organisation level (Fredrickson & Cameron, 1999).	Randomised Cluster Trial Quasi-experimental

4.4 Unique contribution of the study

The unique contribution of the current research is the exploration of the impact of CL on empathy and children's experiences of bullying, building upon limited research in this area.

The study also aimed to replicate and extend previous research by further exploring the impact of CL on PA and exploring the impact on the status hierarchies within the peer ecology. It raised further questions as to whether research should be looking for overall increases in PA or whether the aim should to be to reduce the hierarchy within the classroom.

This work has also highlighted the importance of implementation fidelity in future research and utilising implementation science research.

4.5 Summary and Conclusions

The current study used a quasi-experimental group design to explore the impact of CL on pupils' PA, empathy and experience of bullying in school.

Firstly, the study indicated no overall significant differences between the intervention and control group in all four PA contexts. There were significant differences between the intervention and control group in some individual classes, although these were significant decreases. Possible explanations and potential theoretical or practical underpinnings have been explored. Findings also suggest that CL may intervene with the peer ecology and classroom hierarchies. PA levels might not increase for all pupils, with the potential for a reduction for those pupils with initial high levels of PA and an increase for those with initial low level of PA, in some contexts.

Secondly, the impact of CL on empathy may be explored further with potential development of an alternative to measure which can be used within educational research.

Finally, the study provided some justification for further research to explore the links between CL and bullying, with anecdotal staff data suggesting a decrease and formal data indicating that CL may have at least a stabilising effect on bullying.

Some caution should be taken when interpreting results, particularly as findings were not consistent across individual classes and contexts. When considering effect sizes some small and medium effect sizes are reported. However, these are not all in the direction expected based on theoretical underpinnings and previous research, illustrating both the challenges in applied, small-scale studies, and also the importance of interrogating the theoretical underpinnings to studies. Above all, research to explore the impact of CL across the peer ecology would be of further value. The study does, however, highlight the importance of intervention fidelity and the complexities involved in the implementation of evidence-informed approaches (Gulliford, 2015). Positive outcomes cannot simply be assumed without close attention to context and delivery. Without this, the introduction of evidence-based interventions may carry the risk of negative effects without a secure process to understand and implement interventions.

5 References

Alabekee, Egbulefu, C., Samuel, A., & Osaat, S. D. (2015). Effect of cooperative learning strategy on students learning experience and acheivements in mathematics. *International Journal of Education Learning and Development*, *3*(4), 67–75.

Allison, P. D. (2006). Change Scores as Dependent Variables in Regression Analysis. *Sociological Methodology*, *20*, 93.

Andre, A., Deneuve, P., & Louvet, B. (2011). Cooperative Learning in Physical Education and Acceptance of Students with Learning Disabilities. *Journal of Applied Sport Psychology*, *23*(4), 474–485.

Antil, L. R., Jenkins, J. R., Wayne, S. K., & Vadasy, P. F. (1998). Cooperative Learning: Prevalence, Conceptualizations, and the Relation Between Research and Practice. *American Educational Research Journal*, *35*(3), 419–454.

Armstrong, B., Johnson, D. W., & Balow, B. (1981). Effects of Cooperative vs individualistic Learning Experiences on Interpersonal Attraction between Learning-Disabled and Normal-Progress Elementary School Students. *Contemporary Educational Psychology*, *6*, 102–109.

Arora, C., & Thompson, D. (1987). Defining bullying for the secondary school. *Educaitonal and Child Psychology*, *4*, 110–120.

Artut, P. D. (2009). Experimental Evaluation of the Effects of Cooperative Learning on Kindergarten Children's Mathematics Ability. *International Journal of Educational Research, 48*(6), 370–380.

Baines, E., Blatchford, P., & Kutnick, P. (2003). Changes in grouping practices over primary and secondary school. *International Journal of Educational Research, 39*(1–2), 9–34.

Baker, R. (2004). Co-operative Learning as a Method for Reducing Bullying (Masters Thesis).

Bandura, A. (1971). *Social Learning Theory*. New York: General Learning Press.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how

another person feels versus imagining how you could feel. *Personailty and Social Psychology Review*, *23*(7), 751–758.

Battistich, V., Solomon, D., & Delucchi, K. (1993). Interaction Processes and Student Outcomes in Cooperative Learning Groups. *The Elementary School Journal*, *94*(1), 19–32.

Baumeister, R. F., & Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, *117*(3), 497–529.

Belacchi, C., & Farina, R. (2012). Feeling and thinking of others: Affective and cognitive empathy and emotion comprehension in prosocial/hostile preschoolers. *Aggressive Behavior*, *38*(2), 150–165.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Czajkowski, S. (2004). Enhancing Treatment Fidelity in Health Behavior Change Studies: Best Practices and Recommendations From the NIH Behavior Change Consortium. *Health Psychology*, *23*(5), 443–451.

Bertucci, A., Conte, S., Johnson, D. W., & Johnson, R. T. (2010). The impact of size of cooperative group on achievement, social support, and self-esteem. *The Journal of General Psychology*, *137*(3), 256–272.

Bierman, K. L. (2004). *Peer Rejection: Developmental processes and intervention strategies*. New York:Guliford Press.

Bierman, K. L. (2011). The promise and potential of studying the "invisible hand" of teacher influence on peer relations and student outcomes: A commentary. *Journal of Applied Developmental Psychology*, *32*, 297–303.

Bogdan, R., & Bilken, S. K. (2007). *Qualitative Research for Education: An introduction to Theory and Methods* (5th Ed). London: Boston.

Bosworth, K., Garcia, R., Judkins, M., & Saliba, M. (2018). The impact of leadership involvement in enhancing high school climate and reducing bullying: An exploratory study. *Journal of School Violence*, *17*(3), 354–366.

Brechwald, W. A., & Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on*

Adolescence, *21*(1), 166–179.

Bridgeman, D. (1981). Enhanced Role Taking through cooperative interdependence: A field study. *Child Develpment*, *52*, 1231–1238.

Buchs, C., Filippou, D., Pulfrey, C., & Volpé, Y. (2017). Challenges for cooperative learning implementation: reports from elementary school teachers. *Journal of Education for Teaching*, *433*(3), 296–306.

Buhs, E. S., Ladd, G. W., & Herald, S. L. (2006). Peer exclusion and victimization: Processes that mediate the relation between peer group rejection and children's classroom engagement and achievement? *Journal of Educational Psychology*, *98*(1), 1–13.

Burke Johnson, R., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26.

Calhoon, M. B., Otaiba, S. Al, Cihak, D., King, A., & Avalos, A. (2007). Effects of a Peer-Mediated Program on Reading Skill Acquisition for Two-Way Bilingual First-Grade Classrooms. *Learning Disability Quarterly*, *30*(3), 169.

Callender, A. A., & Mcdaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*, 30–41.

Capodieci, A., Rivetti, T., & Cornoldi, C. (2016). A Cooperative Learning Classroom Intervention for Increasing Peer's Acceptance of Children With ADHD. *Journal of Attention Disorders*.

Capodieci, A., Rivetti, T., & Cornoldi, C. (2016). A Cooperative Learning Classroom Intervention for Increasing Peers Acceptance of Children With ADHD. *Journal of Attention Disorders*,

Carbone-Lopez, K., Esbensen, F. A., & Brick, B. T. (2010). Correlates and consequences of peer victimization: Gender differences in direct and indirect forms of bullying. *Youth Violence and Juvenile Justice*, *8*(4), 332–350.

Carnine, D. (1997). Bridging the reseatch-to-practice gap. *Exceptional Children*, *63*(4), 513–521.

Casey, A., Dyson, B., & Campbell, A. (2009). Action research in physical education: focusing beyond myself through cooperative learning. *Educational Action Research*, *17*(3), 407–423.

Choi, J., Johnson, D. W., & Johnson, R. (2011). Relationships among cooperative learning experiences, social interdependence, children's aggression, victimization, and prosocial behaviors. *Journal of Applied Social Psychology*, *41*(4), 976–1003.

Closson, L. M. (2009). Aggressive and prosocial behaviors within early adolescent friendship cliques: What's status got to do with it?. *Merrill-Palmer Quarterly (1982-)*, 406-435.

Coe, R. (2002). It's the effect size, stupid: what effect size is and why it is important. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London: London : Routledge.

Coie, J. (1990). Toward a theory of peer rejection. In S. R. Asher, & J. D. Coie (Eds.), Peer rejection in childhood. In *In S. R. Asher, & J. D. Coie (Eds.), Peer rejection in childhood* (pp. 365–398). Cambridge: Cambridge University Press.

Coie, J. ., & Dodge, K. . (1998). Aggression and antisocial behavior. In *In N. Eisenberg (Vol. Ed.) & W. Damon (Ed.), Handbook of child psychology, Vol. 3: Social, emotional, and personality development* (5th ed, pp. 779–862). New York: Wiley.

Coie, J., & Dodge, K. (1983). Continuities and Changes in Children's Social Status: A Five Year Longitudinal Study. *Merrill-Palmer Quarterly*, *29*, 261–282.

Cook, B. ., & Odom, S. . (2013). Evidence-Based Practices and Implementation Science in Special Education. *Exceptional Children*, *79*(2), 135–144.

Cowie, H., Smith, P., Boulton, M., & Laver, R. (1994). *Cooperation in the Multi-Ethnic Classroom: The Impact of Cooperative Group Work in Social Relationships in Middle Schools*. Taylor and Francis Publishers.

Craig, J. (2010). *Promoting peer accetance in the classroom: an evaluation of a coopertaive learning intervention in a mainstream primary school*. University of Nottingham.

Cuff, B., Brown, S., Taylor, L., & Howat, D. (2016). Empathy: A Review of the Concept. *Emotion Review*, *8*(2), 144–153.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology : using SPSS for Windows*. (J. Reidy, Ed.) (4th ed.). Harlow : Pearson Prentice Hall.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control?. *Clinical psychology review*, *18*(1), 23-45

Davis, M. . (1980). A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85.

Davison, L., Galbraith, I., & McQueen, M. (2008). Cooperative learning: a partnership between an EPS and a school. *Educational Psychology in Practice*, *24*(4), 307–317.

de Bruyn, E. H., Cillessen, A. H. N., & Wissink, I. B. (2010). Associations of peer acceptance and perceived popularity with bullying and victimization in early adolescence. *Journal of Early Adolescence*, *30*(4), 543–566.

De Rosier, M. ., Cillessen, A. H. ., Coie, J. ., & Dodge, K. . (1994). Group social context and children's aggressive behavior. *Child Development*, *65*, 1068–1079.

Deutsch, M. (1949). A Theory of Co-operation and Competition. *Human Relations* , *2*, 129–152.

Diehl, D. S., Lemerise, E. A., Caverly, S. L., Ramsay, S., & Roberts, J. (1998). Peer Relations and School Adjustment in Ungraded Primary Children. *Journal of Educational Psychology*, *90*(3), 506–515.

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest design. *Work*, *20*, 159–165.

Dishion, T. J., Patterson, G. R., Stoolmiller, M., & Skinner, M. L. (1991). Family, School, and Behavioral Antecedents to Early Adolescent Involvement With Antisocial Peers. *Developmental Psychology*, *27*(1), 172–180.

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., & Price, J. M. (2003). Peer Rejection and Social Information-Processing Factors in the Development of Aggressive Behavior Problems in Children. *Society*, *74*(2),

374–393.

Dovido, J., Gaertner, S., & Saguy, T. (2009). Commonality and the complexity of "we" social attitudes and social change. *Personailty and Social Psychology Review*, *13*, 3–20.

Dugard, P., & Todman, J. (1995). Analysis of Pre-test-Post-test Control Group Designs in Educational Research. *Educational Psychology*, *15*(2), 181–198.

Durlak, J. A. (2010). The importance of doing well in whatever you do: A commentary on the special section, Implementation research in early childhood education. *Early Childhood Research Quarterly*, *25*, 348–357.

Eccles, J. S. (1999). The development of children ages 6 to 14. *Future of Children*.

Eisenberg, N., & Fabes, R. (1990). Empathy: Conceptualization, Measurement and relation to Prosocial Behavior. *Motivation and Emotion* , *14*(2), 131–149.

Elliott, J. (2002). Editorial. Psychological influences upon educational interventions. *Educational and Child Psychology*, *17*(3), 4–5.

Farmer, T. W., Mcauliffe Lines, M., & Hamm, J. V. (2011). Revealing the invisible hand: The role of teachers in children's peer experiences. https://doi.org/10.1016/j.appdev.2011.04.006

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G * Power 3 . 1 : Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4),

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G * Power 3 : A flexible statistical power analysis program for the social , behavioral , and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Faupel, A. (2003). *Emotional Literacy Assessment and Intervention*. London: GL Assessment LTD.

Feshbach N.D. (1984) Empathy, Empathy Training and the Regulation of Aggression in Elementary School Children. In: Kaplan R.M., Konečni V.J., Novaco R.W. (eds) Aggression in Children and Youth. NATO ASI Series (Series D: Behavioural and Social Sciences), vol 17. Springer, Dordrecht.

Fixsen, D., Blase, K., Horner, R., & Sugai, G. (2009). *Concept Paper Developing the Capacity for Scaling Up the Effective Use of Evidence-Based Programs in State Departments of Education*.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature Acknowledgments Recommended citation: For more information*. *Implementation Research: A Synthesis of the Literature*. Tampa, FL: University of South Florida.

Forman, S. G., Shapiro, E. S., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., Stoiber, K. C. (2013). Implementation Science and School Psychology.

Frederickson, J., & Cameron, R. (1999). *Psycholgy in Education Portfolio*. Windsor: NFER-Nelson.

Frederickson, N., & Furnham, A. (1998). *Social Inclusion Survey (SIS). Psychology in Education Portfolio: Social Skills and Emotional Intelligence*. NFER-Nelson.

Frederickson, N., & Furnham, A. (2004). "The relationship between sociometric status and peer assessed behavioural characterstics of included pupils who have moderate learning difficulties and their classroom peers." *British Journal of Educational Psychology, 74*(3), 391–410.

Frey, K. S., Hirschstein, M. K., Edstrom, L. V., & Snell, J. L. (2009). Observed Reductions in School Bullying, Nonbullying Aggression, and Destructive Bystander Behavior: A Longitudinal Evaluation. *Journal of Educational Psychology*, *101*(2), 466–481.

Garandeau, C. F., Lee, I. A., & Salmivalli, C. (2013). Inequality Matters: Classroom Status Hierarchy and Adolescents' Bullying. *Article in Journal of Youth and Adolescence*.

Garton, A. F., & Gringart, E. (2005). *The development of a scale to measure empathy in 8-and 9-year old children 1*. *Australian Journal of Education and Developmental Psychology* (Vol. 5).

Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S., Gillies, J., & Ngeow, E. (2010). Major ingredients of fidelity: A review and scientific guide to improving

quality of intervention research implementation. *Clinical Psychology Review*, *31*, 79–88.

Gendron, B. P., Williams & Nancy, K. R., & Guerra, G. G. (2011). An Analysis of Bullying Among Students Within Schools: Estimating the Effects of Individual Normative Beliefs, Self-Esteem, and School Climate. *Journal of School Violence*, *10*(2), 150–164.

Ghasemi, A., & Zahediasl, S. (2012). Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab*, *10*(2), 486–495.

Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology*, *41*(4), 235–284.

Gillies, R. (2002). The Long-term Effects of Cooperative Learning on Children's Behaviour and Interactions. *Asia Pacific Journal of Education*, *22*(1), 28–37.

Gillies, R. (2002). The Long-term Effects of Cooperative Learning on Children's Behaviour and Interactions. *Asia Pacific Journal of Education*, *22*(1), 28-37.

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, *39*(1–2), 35–49.

Gillies, R. M. (2016a). Cooperative Learning: Review of Research and Practice. *Australian Journal of Teacher Education Australian Journal of Teacher Education Australian Journal of Teacher Education*, *41*(3).

Gillies, R. M. (2016b). Cooperative Learning: Review of Research and Practice. *This Journal Article Is Posted at Research Online. Recommended Citation Gillies, R. M*, *41*(3), 41.

Gillies, R. M., & Ashman, A. F. (2000). The effects of cooperative learning on students with learning difficulties in the lower elementary school. *Journal of Special Education*, *34*(1), 19–27.

Godleski, S. A., Kamper, K. E., Ostrov, J. M., Hart, E. J., & Blakely-McClure, S. J. (2015). Peer victimization and peer rejection during early childhood. *Journal of Clinical Child & Adolescent Psychology*, *44*(3), 380-392.

Gonyea, R. M. (2005). Self-Reported Data in Institutional Research: Review and Recommendations. *New Directions for Institutional Research*, *127*, 73–88.

Goodwin, M. . (1999). Cooperative Learning and social skill: What skills to teach and how to teach them. *Intervention in School and Clinic*, *35*(1), 29–33.

Goudas, M., & Magotsiou, E. (2009). The Effects of a Cooperative Physical Education Program on Students' Social Skills. *Journal of Applied Sport Psychology*, *21*(3), 356–364.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Applied and Practice-Based Research. Special Edition of Research Papers in Education*, *22*(2), 213–228.

Gough, D., Oliver, S., & Thomas, J. (2013). *An Introduction to Systematic Reviews*. London: Sage Publications.

Gulliford, A. (2015). Evidence Based-Practice in Educational Psychology: the Nature of the Evidence. In *In (Eds.) Cline, T. Gulliford, S., amd Birch, S (2015) (Eds.) Educational Psychology*. London: Routledge.

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, *25*(3), 329–347.

Hanewald, R. (2013). Transition Between Primary and Secondary School : Why it is Important and How it can be Supported. *Australian Journal of Teacher Education*, *38*(1), 65–74.

Harris, A. D., Mcgregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The Use and Interpretation of Quasi-Experimental Studies in Medical Informatics. *Journal of American Medical Informatics Associations*, *13*(1), 16–23. 9

Harwell, M. R. (1988). Choosing between parametric and nonparametric tests. *Journal of Counseling & Development*, *67*(1), 35-38.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, *71*(1), 105-131.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analysis relating to acheivement*. London: Routledge.

Hawley, P. H. (1999). The ontogenesis of social dominance: A strategy- based evolutionary perspective. *Developmental Review*, *19*, 97–132.

Hendrickx, M. M. H. G., Mainhard, M. T., Boor-Klip, H. J., Cillessen, A. H. M., & Brekelmans, M. (2015). Social dynamics in the classroom: Teacher support and conflict and the peer ecology. *Teaching and Teacher Education, 53*, 30–40.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, *2*(3), 172-177.

Hong, J. S., & Espelage, D. L. (2012). A review of research on bullying and peer victimization in school: An ecological system analysis. *Aggression and Violent Behavior*, *17*(4), 311–322.

Hughes, J. N., Cavell, T. A., & Willson, V. (2001). Further support for the developmental significance of the quality of the teacher–student relationship. *Journal of school psychology*, *39*(4), 289-301.

Hurry, J., & Doctor, E. (2007). Assessing literacy in children and adolescents. *Child and Adolescent Mental Health*, *12*(1), 38–45.

Hymel, S., McClure, R., Miller, M., Shumka, E., & Trach, J. (2015). Addressing school bullying: Insights from theories of group processes. *Journal of Applied Developmental Psychology*, *37*(1), 16–24.

Jacques, N., Wilton, K., & Townsend, M. (1998a). Cooperative learning and social acceptance of children with mild intellectual disability. *Journal of Intellectual Disability Research*, *42*(1), 29–36.

Jacques, N., Wilton, K., & Townsend, M. (1998b). Cooperative learning and social acceptance of children with mild intellectual disability. *Journal of Intellectual Disability Research*, *42*(1), 29–36.

Johnson, David W.; Johnson, Robert, T. (1991). Cooperative learning and classroom and school climate. In *Fraser, B. J., & Walberg, H. J. (Eds.), Educational environments: Evaluation, antecedents and consequences* (pp. 55–74). Elmsford,

NY, US: Pergamon Press.

Johnson, D. ., & Johnson, R. . (2002). Learning Together and Alone: Overview and Meta-analysis. *Asia Pacific Journal of Education, 22-1, 95-105*, *22*(1), 95–105.

Johnson, D. ., Johnson, R. ., & Johnson Holubec, E. (1991). *Cooperation in the Classroom*. Edina, Minnesota: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2005). New developments in social interdependence theory. *Genetic, social, and general psychology monographs*, *131*(4), 285-358

Johnson, D. W., Johnson, R., & Anderson, D. (1983). Classroom interdependence and classroom climate. *The Journal of Psychology*, *114*(1), 135–142.

Johnson, D. W., & Johnson, R. T. (1989). Social Skills for Successful Group Work. *Educational Leadership*, *47*(4), 29–34.

Johnson, D. W., & Johnson, R. T. (2009). An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, *38*(5), 365–379.

Johnson, D. W., & Johnson, R. T. (2012). Restorative justice in the classroom: Necessary roles of cooperative context, constructive conflict, and civic values. *Negotiation and Conflict Management Research*, *5*(1), 4–28.

Johnson, D. W., Johnson, R. T., Roseth, C., & Shin, T. S. (2014). The relationship between motivation and achievement in interdependent situations. *Journal of Applied Social Psychology*, *44*(9), 622–633.

Johnson, D. W., Johnson, R. T., Smith, K. A., & Smith, K. (2013). *Cooperative Learning: Improving University Instruction By Basing Practice On Validated Theory Journal on Excellence in University Teaching Cooperative Learning: Improving University Instruction By Basing Practice On Validated Theory*.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). Cooperative Learning Methods: A Meta-Analysis.

Johnson, D. W., Johnson, R. T., & Taylor, B. (1993). Impact of cooperative and individualistic learning on high-ability students' achievement, self-esteem, and

social acceptance. *Journal of Social Psychology*, *133*(6), 839–844.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). The effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin*, *89*, 47–62.

Johnson, R., & Johnson, D. (1992). An overview of cooperative learning. *Creative and Collabortaive Learning*, *1*(2), 1–21.

Jolliffe, D., & Farrington, D. P. (2006). Examining the Relationship Between Low Empathy and Bullying. *Agressive Behavior*, *32*, 540–550.

Jolliffe, W., & Snaith, J. (2017). Developing cooperative learning in initial teacher education: indicators for implementation. *Journal of Education for Teaching*, *43*(3), 307–315.

Joyce, B., & Showers, B. (2002). *Student Achievement through Staff Development* (3rd ed.). London: Longman.

Jules. (1991). Interaction Dynamics of Cooperative Learning Groups in Trinidad's Secondary Schools. *Adolescence*, *26*(104), 931–949.

Kagan, S., & Kagan, M. (2015). *Kagan Cooperative Learning*. San Clemente, CA: Kagan Publishing.

Kagan, S. (1989). The Structural Approach to Cooperative Learning. *Educational Leadership*, *47*(4), 12–15.

Kelly, B., & Perkins, D. . (2012). *Handbook of implementation science for psyhology in education*. Cambridge: University Press.

Knapp, T. R., & Schafer, W. D. (2009). *From Gain Score t to ANCOVA F (and vice versa)* (Vol. 14).

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*.

Larson, C. O., Dansereau, D. F., O'Donnell, A. M., Lambiotte, J. G., & Rocklin, A. R. (1985). Effects of Metacognitive and Elaborative Activity on Cooperative Learning and Transfer. *Contemporary Educational Psychology*, *10*, 342–448.

Lewin, K. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & D'Apollonia, S. (1996). Within-Class Grouping: A Meta-Analysis. *Review of Educational Research*.

Lyon, A. R., Cook, C. R., Brown, E. C., Locke, J., Davis, C., Ehrhart, M., & Aarons, G. A. (2018). Assessing organizational implementation context in the education sector: confirmatory factor analysis of measures of implementation leadership, climate, and citizenship. *Implementation Science*, *12*(5), 1–14.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in educational research*, *16*(2), 193-205.

Malti, T., Perren, S., & Buchmann, M. (2010). Children's peer victimization, empathy, and emotional symptoms. *Child Psychiatry and Human Development*, *41*(1), 98–113.

Marr, M. B. (1997). Cooperative learning: A brief Overview. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(1), 7–20.

Marsden, E., & Torgerson, C. J. (2012). Oxford Review of Education Single group, pre-and post-test research designs: Some methodological concerns Single group, pre-and post-test research designs: Some methodological concerns.

Mashender. (1986). *Lets Cooperate*. London: Pearce Education Project.

Maslow, A. (1970). *Motivation and Personality* (2nd ed.). New York: Harper & Row.

Mendive, S., Weiland, C., & Snow, C. (2016). Opening the Black Box : Intervention Fidelity in a Randomized Trial of a Preschool Teacher Professional Development Program, *108*(1), 130–145.

Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How Effective Are School Bullying Intervention Programs? A Meta-Analysis of Intervention Research. *School Psychology Quarterly*, *23*(1), 26–42.

Mertens, D., & McLaughlin, J. (1995). *Reseach Methods in Special Education*. London: Thousand Oaks, California.

Mertens, D., & McLaughlin, J. (2004). Research and Evaluation Methods in Special

Education. Thousand Oaks, California.

Moir, T. (2018). Why Is Implementation Science Important for Intervention Design and Evaluation Within Educational Settings? *Frontiers in Education*, *3*(July), 1–9. https://doi.org/10.3389/feduc.2018.00061

Murphy, E., Grey, I. M., & Honan, R. (2005). Co-operative learning for students with difficulties in learning: a description of models and guidelines for implementation. *British Journal of Special Education*, *32*(3), 157–164.

Murphy, E., Grey, I. M., & Honan, R. (2005). Co-operative learning for students with difficulties in learning: a description of models and guidelines for implementation. *British Journal of Special Education*, *32*(3), 157–164.

Murphy, S. L., & Gutman, S. A. (2012). Intervention fidelity: A necessary aspect of intervention effectiveness studies. *American Journal of Occupational Therapy*, *66*(4), 387–388.

Oakes, J. M., & Feldman, H. A. (2001). Statistical Power for Non equivalent Pretest-Posttest Designs: The Impact of Change-Score Versus ANCOVA Models. *Evaluation Review*, *25*(1), 2–28.

Olweus, D. (1993). *Bullying at School*. Cambridge: MA: Blackwell.

Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the olweus bullying prevention program. *American Journal of Orthopsychiatry*, *80*(1), 124–134.

Oortwijn, M. B., Boekaerts, M., Vedder, P., & Fortuin, J. (2008). The impact of a cooperative learning experience on pupils' popularity, non-cooperativeness, and interethnic bias in multiethnic elementary schools. *Educational Psychology*, *28*(2), 211–221.

Oortwijn, M. B., Boekaerts, M., Vedder, P., & Strijbos, J. W. (2008). Helping behaviour during cooperative learning and learning gains: The role of the teacher and of pupils' prior knowledge and ethnic background. *Learning and Instruction*, *18*(2), 146–159.

Osborne, J. W. (2014). The power of outliers and fringeliers (and why researchers should always check for them). *Practical Asessment, Research and Evaluation*,

9(6), 1–8.

Paciello, M., Fida, R., Cerniglia, L., Tramontano, C., & Cole, E. (2013). High cost helping scenario: The role of empathy, prosocial reasoning and moral disengagement on helping behavior. *Personality and Individual Differences*, *55*(1), 3-7.

Pellegrini, A. D., & Bartini, M. (2000). A Longitudinal Study of Bullying, Victimization, and Peer Affiliation During the Transition From Primary School to Middle School. *American Educational Research Journal*, *37*(3), 699–725.

Pellegrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British journal of developmental psychology*, *20*(2), 259-280.

Perestelo-Perez, L. (2013). Standards on how to develop and report systematic reviews in psychology and health. *International Journal of Clinical and Health Psychology*, *13*(1), 49–57.

Peterson, R., Johnson, D., & Johnson, R. (1991). Effects of cooperative learning on perceived status of male and female pupils. *The Journal of Social Psychology*, *131*(5), 717–735.

Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. (Malden, Ed.). Oxford.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Piaget, J. (1926). *The Language and Thought of the Child*. New York: Harcourt Brace.

Piaget, J. (1965). *The moral judgement of the child*. New York: Free Press.

Piercy, M., Wilton, K., & Townsend, M. (2002). Promoting the social acceptance of young children with moderate-severe intellectual disabilities using cooperative learning techniques. *American Journal on Mental Retardation, 107*(5), 352–360.

Piercy, M., Wilton, K., & Townsend, M. (2002). Promoting the Social Acceptance of Young Children With Moderate–Severe Intellectual Disabilities Using Cooperative-Learning Techniques. *American Journal on Mental Retardation*,

107(5), 352–360.

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of educational psychology*, *105*(4), 1050.

Putnam, J., Markovchick, K., Johnson, D. W., & Johnson, R. T. (1996). Cooperative learning and peer acceptance of students with learning disabilities. *The Journal of Social Psychology*, *136*(6), 741–752.

Quinn, M. M. (2002). Changing Antisocial Behavior Patterns in Young Boys: A Structured Cooperative Learning Approach. *Education and Treatment of Children*, *25*(4), 380–395.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, *2*(1), 21-33.

Reid, C., Davis, H., Horlin, C., Anderson, M., Baughman, N., & Campbell, C. (2013). The Kids' Empathic Development Scale (KEDS): A multi-dimensional measure of empathy in primary school-aged children. *British Journal of Developmental Psychology*, *31*(2), 231–256.

Roberts, W., Strayer, J., & Denham, S. (2014). Empathy, anger, guilt: Emotions and prosocial behaviour. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *46*(4), 465.

Robinson, A. (1990). Cooperation or Exploitation? the Argument against Cooperative Learning for Talented Students. *Journal for the Education of the Gifted*, *14*(1), 9–27.

Robson, C., & McCartan, K. (2016). *Real world research : a resource for users of social research methods in applied settings*. (K. McCartan 1980-, Ed.) (4th ed.). Chichester: Chichester : Wiley.

Rodkin, P. C., & Gest, S. D. (2010). Teaching practices, classroom peer ecologies, and bullying behaviors among schoolchildren. In *Bullying in North American schools* (pp. 95-110). Routledge.

Roland, E., & Galloway, D. (2002). Classroom influences on bullying. *Educational research*, *44*(3), 299-312.

Roseth, C., Johnson, D., & Johnson, R. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goa structures. *Psychological Bulletin., 134*(2), 223–246.

Rowe, F., & Stewart, D. (2009). Promoting connectedness through whole-school approaches: A qualitative study. *Health Education*, *109*(5), 396-413.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78.

Schäfer, M., Korn, S., Brodbeck, F. C., Wolke, D., & Schulz, H. (2004). Bullying roles in changing contexts: The stability of victim and bully roles from primary to secondary school. *International Journal of Behavioral Development*, *29*(4), 323-335.

Schäfer, M., Korn, S., Brodbeck, F. C., Wolke, D., & Schulz, H. (2005). Bullying roles in changing contexts: The stability of victim and bully roles from primary to secondary school. *International Journal of Behavioral Development*, *29*(4), 323–335.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in Treatment Integrity Research: Multidisciplinary Perspectives on the Conceptualisation, Measurement, and Ehancement of Treatment Integrity. *School Psychology Review*, *38*(4), 460–475.

Scotland, J. (2012). Exploring the Philosophical Underpinnings of Research: Relating Ontology and Epistemology to the Methodology and Methods of the Scientific, Interpretive, and Critical Research Paradigms. *English Language Teaching*, *5*(9).

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002a). *Experimental and quasi-experimental designs for generalized causal inference*. (T. D. Cook & D. T. Campbell 1916-1996, Eds.). Belmont, Calif.: Belmont, Calif. : Wadsworth Cengage Learning.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002b). Experimental and Quasi-

Experimental for Generalized Designs Causal Inference. *Handbook of Industrial and Organizational Psychology*, *223*, 623.

Sharan, Y. (2010). Cooperative learning for academic and social gains: Valued pedagogy, problematic practice. *European Journal of Education*, *45*(2), 300–313.

Shwalb, B. J., & Shwalb, D. W. (1995). Cooperative learning in cultural contexts: an integrative review. *International Journal of educational research*, *23*(3), 293-300.

Skinner, B. (1953). *Science and Human Behaviour*. New York: Macmillan.

Skon, L., Johnson, D. W., & Johnson, R. T. (1981). Cooperative peer interaction versus individual competition and individualistic efforts: Effects on the acquisition of cognitive reasoning strategies. *Journal of Educational Psychology*, *73*(1), 83–92.

Slavin, R. E. (1991). Student Team Learning: A Practical Guide to Cooperative Learning. Washington DC: National Education Association

Slavin, R. (2004). When and why does coopertaive learning increase achievement. In *In Daniels, H & Edwards, A, The Routledge Falmer Reading in Psychology of Education* (pp. 271–293). London: Routledge Falmer.

Slavin, R. E. (1990). Research on cooperative learning: Consensus and controversy. *Educational Leadership*, *47*(4), 52–54.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary educational psychology*, *21*(1), 43-69.

Slavin, R. E. (2014). Cooperative Learning and Academic Achievement: Why Does Groupwork Work?, Annals of Psychology, *30*(3), 785–791.

Slavin, R. E. (2015). Cooperative learning in elementary schools. *Education 3-13*, *43*(1), 5–14.

Smith, H. (2017). *The impact of Cooperative Learning on pupils Peer Acceptance and Perceptions of Peer Support*. University of Nottingham.

Smith, P. K., Boulton, M. J., & Cowie, H. (1993). The impact of cooperative group work on ethnic relations in middle school. *School Psychology*

International, *14*(1), 21-42.

Stahl, R. J. (1994). *The Essential Elements of Cooperative Learning in the Classroom. ERIC Digest*.

Stephen, T. T., Kreskey, D., Bolling, M., Psimas, L., Graybill, E., Albritton, K., & Schwartz, A. (2012). Creating consultee change: A theory-based approach to learning and behavioral change processes in school-based consultation. *Consulting Psychology Journal*, *64*(1), 63–82.

Stephen, W., & Finlay, K. (1999). Stephan-Finlay.-1999.-The-Role-of-Empathy-in-Improving-Intergroup-Relations. *Journal of Social Issues*, *55*(4), 729–743.

Stietz, J., Jauk, E., & Krach, S. (2019). Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Frontiers in Psychiatry*, *1*, 126.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education*, *4*(3), 279-282.

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying? Linking research to educational practice. *Educational researcher*, *39*(1), 38-47.

Tran, V. D. (2013). Theoretical Perspectives Underlying the Application of Cooperative Learning in Classrooms. *International Journal of Higher Education*, *2*(4), 101-115.

Terwel, J., Gilles, R., van den Eeden, P., & Hoek, D. (1999). Opening the black box: Processes in cooperative learning from a longitudinal multilevel perspective. In *Paper presented at the American Educational Research Associations Annual Meeting*. Montreal.

Topping, K., Thurston, A., Tolmie, A., Christie, D., Murray, P., & Karagiannidou, E. (2011a). Cooperative learning in science: intervention in the secondary school. *Research in Science & Technological Education*, *29*(1), 91–106.

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, *7*(1), 27-56.

Unnever, J. D., & Cornell, D. (2003). The Culture of Bullying in Middle School School Climate. *Journal of School Violence*, *2*, 5–27.

Vaillancourt, T., Hymel, S., & Mcdougall, P. (2001). Bullying Is Power: Implications for School-Based Intervention Strategies. *Journal of Applied School Psychology 19*(2), 157–176.

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*(9), 920–925.

Van De Mortel, T. F. (2008). *Faking it: social desirability response bias in self-report research*. *Australian Journal of Advanced Nursing*, 24, 4, 40-48.

Van Ryzin, M. J., & Roseth, C. J. (2018). Cooperative learning in middle school: A means to improve peer relations and reduce victimization, bullying, and related outcomes. *Journal of Educational Psychology.*

Vasileiadou, M. (2009). Cooperative learning and its effects on pre-primary, marginalized children. *Emotional and Behavioural Difficulties*, *14*(4), 337–347.

Vaughn, J. B. (1994). Cooperative learning and young children: Emerging cooperative behaviors. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *54*.

Vreeman, R., & Carroll, A. (2007). A Systematic Review of School-Based Interventions to Prevent Bullying. *American Medical Association*, *161*, 78–88.

Vygotsky, L. . (1978). *Mind in Society*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard Press.

Wang, C., Berry, B., & Swearer, S. M. (2013). The Critical Role of School Climate in Effective Bullying Prevention, *52*, 296–302.

Warden, D., & Mackinnon, S. (2003). Prosocial children, bullies and victims: An investigation of their sociometric status, empathy and social problem-solving strategies. *British Journal of Developmental Psychology*, *21*(3), 367-385.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, *79*, 1–28.

Webb, N. M., & Farivar, S. (1994). Promoting Helping Behavior in Cooperative Small Groups in Middle School Mathematics. *American Educational Research Journal*, *31*(2), 369–395.

Williams-Jones, T. (2004). An investigation of the use of cooperative learning techniques with a sample of children (0--4) across traditional daycare and playgroup learning communities *(Doctoral Dissertation).*

Williams, A., O'Driscoll, K., & Moore, C. (2014). The influence of empathic concern on prosocial behavior in children. *Frontiers in Psychology*, *5*, 425. https://doi.org/10.3389/fpsyg.2014.00425

Winter, G. (2000). The Qualitative Report A Comparative Discussion of the Notion of "Validity" in Qualitative and Quantitative Research. *The Qualitative Report*, *4*(3), 1–14.

Wright, D. . (2003). Making friends with your data: Improving how statistics are conducted and reported. *British Journal of Educational Psychology*, *73*(1), 123–136.

Yee Mikami, A., Boucher, M. A., & Humphreys, K. (2005). Prevention of Peer Rejection Through a Classroom-Level Intervention in Middle School. *The Journal of Primary Prevention*, *26*(1), 5-23.

Zakaria, E., Solfiri, T., Daud, Y., & Abidin, Z. Z. (2013). Effect of Cooperative Learning on Secondary School Students' Mathematics Achievement. *Creative Education*, *4*(2), 98–100.

6 Appendices

Appendix A A table depicting the source and number of results obtained from the systematic search

Source	Number of results screened	Inclusion/exclusion applied to title/ abstract	Included Articles		Full Text obtained
EBSCO Host	51	15 4 were excluded because they were duplicates. The remaining were excluded for not being related to Cooperative Learning, not being related to the variables and not being experimental.	*1.Ng (1999)* *2. Bell (1997)* *3. Asilidae (2009)* *4. Artut (2009)* *5. Slavin (1983)* *6. Putnam (1993)* *7. Fetsch (2002)* *8. Piercy (2002)*	*9. Crockenberg (1977)* *10. Hertz-Lazarowitz (1980)* *11. Watson (1988)* *12. Madden (1981)* *13. Johnson (1993)* *14. Madden (1983)* *15. Singh (2010)*	***6*** *1. Ng (1991)* *2. Vasileiadou (2009.* *3. Artut (2009)* *4. Putnam (1993)* *5 Piercy (2002)* *6. Hertz-Lazarowitz (1980)*
Psych Info	35	12 7 were removed as they were duplicates. The remaining were removed for not being related to variables and not being experimental	1. *Vasileiadou (2009)* 2. *Piercy (2002)* 3. *Jacques (1998)* 4. *Crockenburg (1979)* 5. *Putnam (1996)* 6. *Del Rio (2017)*	7. *Ryzin (2018)* 8. *Leon (2012)* 9. *Alexander (2008)* 10. *Choi (2011)* 11. *Weitzman (1994)* 12. *Sanders (1991)*	***6*** *1 article was removed as it was in Spanish* *4 were excluded as no full text was available* *1 because it was qualitative*

NU Search	56	9 38 were removed for the following reasons: • duplications of same article • not related to three variables (i.e. related to academic achievement) • Not experimental I.e. case study	1. Choi et al (2011) 2. Capideci et al (2016) 3. Jacques et al (1998) 4. Quinn (2002) 5. Piercy (2002) 6. Vasileidou (2009)	7. Gillies (2002) 8. Cowie & Berdondini (2001) 9. Ryzin (2018) 10. Oortwign (2008)	10 full text articles obtained
Total remaining	142	36	37		22 11 were removed as they were duplicates across searches

Appendix B Summary of Full text studies excluded from synthesis

Reference	Reason for Exclusion
Artut, P. D. (2009). Experimental evaluation of the effects of cooperative learning on kindergarten children's mathematics ability. International journal of educational research, 48(6), 370-380.	Related to maths ability, no measures relating to one of the variables
Choi, J., Johnson, D. W., & Johnson, R. (2011). Relationships among cooperative learning experiences, social interdependence, children's aggression, victimization, and prosocial behaviors. *Journal of Applied Social Psychology*, *41*(4), 976-1003.	Not an experimental study
Cowie, H., and L. Berdondini. "Children's reactions to cooperative group work: a strategy for enhancing peer relationships among bullies, victims and bystanders." *Learning and instruction* 11, no. 6 (2001): 517-530.	Referred to "Cooperative Group Work" which is not considered the same as CL.
Gillies, R. (2002). The Long-term Effects of Cooperative Learning on Children's Behaviour and Interactions. *Asia Pacific Journal of Education*, *22*(1), 28-37.	Approach not considered to be synonymous with CL

Hertz-Lazarowitz, R., Sharan, S., & Steinberg, R. (1980). Classroom learning style and cooperative behavior of elementary school children. *Journal of Educational Psychology*, *72*(1), 99.	Approach not considered to be synonymous with CL
Ng, M., & Lee, C. K. E. (1999). The Effects of Cooperative Learning on the Cross-Ethnic Friendship Choices of Children.	Approach not considered to be synonymous with CL

Appendix C A table summarising the Weight of Evidence Criteria

WoE A Methodological Quality	**Low**	**Medium**	**High**
1. Hypotheses /aims/ objectives are clearly described 2. Main outcomes to be measures are clearly described in the introduction or methods section 3. Participants characteristics are clearly described (e.g. number, age, gender). 4. The Cooperative learning intervention is clearly described (including duration, frequency, facilitator and approach/model used for Cooperative Learning (e.g. Kagan etc). 5. The fidelity of the intervention is cleary described 6. Random assignment to conditions 7. Methods and procedures are clearly described 8. Control group and/or comparison group 9. Reliability/Validity of relevant measures described 10. Main findings are described 11. Exact p levels are reported 12. Limitations are discussed	Fulfils 0-6 of the criteria	Fulfils 6-9 of the criteria	Fulfils 9-12 of the criteria
WoE B: Methodological appropriateness	**Low**	**Medium**	**High**
1. Randomised Control Trial (RCT) or quasi-experimental design 2. Random allocation to groups 3. Participants appropriate (e.g. age, setting)	Fulfils 0 -3 of the criteria	Fulfils 3-6 of the criteria	Fulfils 6-9 of the criteria

4. Control or comparison group 5. Measures used are related to the research question 6. Pre and post data collected 7. Social Skills Training is present 8. Teacher selected Heterogenous groups 9. Method of CL is clearly stated			

WoE C: Relevance of Evidence	**Low**	**Medium**	**High**
1. Study was based in the UK 2. Cooperative learning as a while class intervention 3. CL delivered by class teachers 4. Participants are predominately in KS2 5. Participants are in a mainstream setting 6. CL used within an academic lesson and within usual curriculum 7. Study evaluates the impact of CL on PA and/or bullying and/or empathy 8. Usual teaching as a control group 9. Participants are in their usual classes	Fulfils 0 -3 of the criteria	Fulfils 3-6 of the criteria	Fulfils 6-9 of the criteria

WoE D was calculated based on an average of A, B and C or the most frequently occurred rating if an average was not possible.

Appendix D A table illustrating a list of studies included and excluded from the unpublished articles systematic review search

Study	**Included/Excluded**	**Reasons**
(Vaughn, 1994)	Excluded	Not related to any of the three variables
(Williams-Jones, 2004)	Excluded	Not related to any of the three variables
Bjerklie (1994)	Excluded	Not related to any of the three variables
Li (2001)	Excluded	Not related to any of the three variables
Wang (2009)	Excluded	Not related to any of the three variables
Broner (1976)	Excluded	Not related to any of the three variables
Stewart (1998)	Excluded	Not related to any of the three variables
Delgado (1987)	Excluded	Not related to any of the three variables
Schlein (1982)	Excluded	Not related to any of the three variables
Edfelt	Excluded	Not related to any of the three variables
Tao (2009)	Excluded	Not related to any of the three variables
Nowak (1996)	Excluded	Not related to any of the three variables
Hightower (1998)	Excluded	Not related to any of the three variables
Haynes (2002)	Excluded	Not related to any of the three variables
Slack (1979)	Excluded	Not related to any of the three variables

Balfry (2009)	Excluded	Not related to any of the three variables
Smith (2018)	Included	Related to CL and PA
Craig (2010)	Included	Related to CL and PA
Hoffman (2010)	Excluded	Not an experimental design
Alexander (2008)	Excluded	Not an experimental design
Sanders (1991)	Excluded	Related to CL and empathy but full text could not be obtained.
Dixon (2006)	Excluded	Not related to CL

Appendix E Information provided to school for research recruitment

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

INSERT DATE

Dear XXXX

I am a second-year Trainee Educational Psychologist at the University of Nottingham, on placement in Staffordshire Educational Psychology Service working within the Lichfield/Cannock District.

As part of my training, I am undertaking research into the impact of **Cooperative Learning** in primary schools. Cooperative Learning is a structured learning approach, using group activities in the existing academic curriculum with the aim of promoting learning and enhancing children's social interaction skills.

I am looking for primary schools with two or more **class teachers in Year 4 and/or Year 5 who would like to learn about Cooperative Learning approaches** and would be willing to incorporate the techniques into their classroom in the 2018 Autumn Term. These approaches can be incorporated into the existing curriculum and into existing lesson plans.

Aim of the Research

The aim of this study is to evaluate the impact of a Cooperative Learning intervention on Peer Acceptance, empathy and pupils' experience of bullying (see below) in school. The intervention, which will run over 11 weeks in the Autumn Term and will involve Cooperative Learning activities being incorporated within daily lessons.

I have attached an information sheet, giving more details about Cooperative Learning.

What are the benefits of taking part in the Study?

For Pupils:

Extensive research already shows Coopertaive Learning has several benefits on:

- Academic Achievement
- Self-Esteem
- Social skills
- Social Inclusion/Peer Acceptance

(Kagan and Kagan, 2015)

For Teachers and schools:

- **Free training and professional development** for the class teachers involved
- Ongoing support, coaching and monitoring during the research period
- A chance for your school to be involved in research
- Impact upon future practice

What do I need to be involved in this opportunity?

To be involved in the research you will need teachers:

- who are teaching in Years 4 or 5 starting from September 2018.
- who are available for a training session in July 2018 and a refresher session in September 2018.
- who are willing to commit to implementing Cooperative Learning techniques into their classroom during the Autumn Term.
- who are happy for me to complete brief questionnaires with their pupils before and after the intervention.

The research requires a "control group" this means there **needs to be at least one parallel year group clas**s. The class teacher of this class will not need to implement the Cooperative Learning approaches and should continue with their usual teaching but needs to be willing for their class to complete some brief questionnaires and for me to visit and carry out observations of group work. The class teacher of the comparison group will be offered the same Cooperative Learning training following the intervention to ensure the control group have access to the same activities.

What about Ethics?

The project will be conducted within the ethical parameters set by the British Psychological Society (BPS) and will have received ethical approval from the University of Nottingham Ethics Committee.

What happens after the research?

Following the research, I would hope to discuss the findings with you and will offer a final report and opportunity for further discussion.

What Next?

If you are interested in being involved in this project, please contact me on the email address below. I will then contact you and can arrange a suitable time to meet with you, or relevant person(s) in school to discuss the project and answer any questions you may have. Schools will be contacted on a first come first serve basis.

At this stage this I am asking for expressions of interest and you are not committed to take part in the research. If you have any queries, please do not hesitate to contact me.

Yours Faithfully,

Rachel Faizey

Trainee Educational Psychologist

XXXXXX Educational Psychology Service

Rachel.Faizey@nottingham.ac.uk

Information for schools

What is Cooperative Learning?

Cooperative Learning is a way of structuring parts of a usual lesson to encourage social interactions, whilst providing group members with the opportunity to practice their social skills.

*"small groups of learners working together as a team to solve a problem, complete a task or accomplish a common goal" (*Artz and Newman, 1990, p.448)

As an example of Cooperative Learning: children might be asked to discuss their answers or ideas with a partner, and then they both share their ideas with another pair of children.

Cooperative Learning approaches provide a structure for group activities which aim to:

- encourage **positive interdependence.** Group members need each other to complete the task
- promote **individual accountability**. Everybody has a stake in the outcome, no freeloading!
- ensure **equal participation.** Tasks are structured so that everyone can make an equal contribution
- actively teach relevant small group **social skills** through the academic curriculum

Cooperative Learning approaches can be easily fitted into the existing curriculum and existing lesson plans.

What does the research show?

Academic skills:

- increased achievement across subjects, age groups and ability levels
- enhanced problem-solving increased time spent on task

Social and emotional development:

- enhanced self esteem
- increased motivation
- greater PA and pupil friendships
- increased acceptance of diversity
- higher behaviour ratings from staff
- enhanced liking of school

Source: Kagan and Kagan (2015)

Appendix F Information Sheet for Class Teachers

School of Psychology

Information Sheet

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Title of Project**:** An evaluation of a cooperative learning intervention on PA, empathy and pupils experience of school

Insert Ethics Approval Number or Taught Project Archive Number: S1078

Researchers: Rachel Faizey (rachel.faizey@nottingham.ac.uk)

Supervisor: Anthea Gulliford (anthea.gulliford@nottingham.ac.uk)

This is an invitation to take part in a research study on **Cooperative Learning.**

Before you decide if you wish to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

Aim of the Research Project

The aim of the research is to evaluate of the impact of a Cooperative Learning intervention on Peer Acceptance, empathy and pupils experience of bullying in school.

To do this, I aim to introduce Cooperative Learning on a whole-class basis in several schools during the 2018 Autumn Term. This will involve social skills training sessions for all the children involved to familiarise them with the skills necessary to participate in Cooperative Learning activities successfully such as listening skills and information sharing. Following this the intervention will run over 10 weeks in summer term and will involve Cooperative Learning activities being incorporated within daily lessons.

What the Research Project entail?

The Research will have an "intervention group" and a "Comparison group" If you are selected to participate as part of the intervention group you will be asked to:

- Attend a training session focussing on social skills and Cooperative Learning structures.
- Incorporate one week of social skills training into the curriculum
- Incorporate Cooperative Learning structures into existing lesson plans a minimum of four times a week as part of a whole class lesson at least four times a week over a 10-week period over the summer term.
- Be available and willing to be part of ongoing monitoring and support visits
- Allow the researcher to collect measures using a short questionnaire with all pupils before and after the intervention period.

- Keep a daily record of the Cooperative Learning sessions and any pupil absence.

If you are selected to take part in Comparison condition you will be asked:

- To make no changes to the teacher methods used within the classroom during research
- To allow the researcher to collect measures using short questionnaires with all pupils before and after the intervention period.
- To be available and willing to be part of ongoing monitoring and support visits

If you are part of the Comparison condition you will be offered the same training following the intervention.

The intervention will last 11 weeks during the summer term.

Participation in this study is totally voluntary and you are under no obligation to take part. You are free to withdraw at any point before or during the study. All data collected will be kept confidential and used for research purposes only. It will be stored in compliance with the Data Protection Act. Pupils are also free to withdraw from taking part in the research at any time and will be reminded of this before and after completing questionnaires.

If you have any questions or concerns, please don't hesitate to ask. We can also be contacted after your participation at the above address.

If you have any complaints about the study, please contact:
Stephen Jackson (Chair of Ethics Committee)
stephen.jackson@nottingham.ac.uk

School of Psychology

Title of Project An evaluation of a cooperative learning intervention on PA, empathy and pupils experience of school

Ethics Approval Number or Taught Project Archive Number:

Researcher(s): Rachel Faizey (rachel.faizey@nottingham.ac.uk)

Supervisor(s): Anthea Gulliford (anthea.gulliford@nottingham.ac.uk)

The participant should answer these questions independently:

- Have you read and understood the Information Sheet?
  YES/NO
- Have you had the opportunity to ask questions about the study?
  YES/NO
- Have all your questions been answered satisfactorily (if applicable)?
  YES/NO
- Do you understand that you are free to withdraw from the study?
  YES/NO
  (at any time and without giving a reason)
- I give permission for my data from this study to be shared with other researchers provided that my anonymity is completely protected. YES/NO
- Do you agree to take part in the study?
  YES/NO

"This study has been explained to me to my satisfaction, and I agree to take part. I understand that I am free to withdraw at any time."

Signature of the Participant: Date:

Name (in block capitals)

I have explained the study to the above participant, and he/she has agreed to take part.

Signature of researcher: Date:

Appendix G Parental Information Sheet and consent forms

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Dear Parent/Carer.

Your child's school has agreed to be involved in a **Cooperative Learning Project** evaluating the effectiveness of cooperative learning on Peer Acceptance, empathy and experience of friendly and bullying behaviour in school.

As part of my doctoral training as a Trainee Educational Psychologist, I am looking at the impact of Cooperative Learning on Peer Acceptance, empathy and pupils' experiences of school in a primary school.

Cooperative Learning is a group intervention approach in which students work together to achieve shared learning goals. Coopertaive Learning provides a way of structuring group activities within the existing curriculum and aims to promote learning and enhance social skills. Cooperative Learning has been found to have several benefits for pupils including on motivation, academic achievement, Peer Acceptance and self-esteem.

Your child will either be part of one of two groups. One group will receive some lessons using a Cooperative Learning format. An example of Cooperative Learning is pupils working together to reach an agreement to answer on a problem or question presented by the class teacher. Each pupil is given a number and any number may be called to share the team's answer.

The other group will receive usual teaching with no addition of cooperative learning activities; however, they will be given access to Cooperative Learning activities upon completion of the study.

I am writing to request consent for your child to complete questionnaires in the Classroom (on two occasions) which will be used to evaluate how useful Cooperative Learning has been. Your child will be asked to complete three questionnaires:

1. The first questionnaire will measure **PA** within the classroom and will ask pupils how much they like to play and work with other children in the class.

2. The second questionnaire will measure **empathy** asking pupils to look at picture-based scenarios and infer what emotions that person or persons in the image might be feeling, and attributing reasons for these feelings.

3. The third questionnaire will measure **pupils' experiences of bullying in school.** Pupils will be asked to fill in a questionnaire which includes statements which describe positive and negative events including statements describing friendly or bullying behaviours. Pupils are asked to indicate whether they had experienced the situation "not at all", "only once" or "more than once"

Children's responses will be kept confidential and will not be shared with school staff or pupils. The results will then be written up as part of my research project. Involvement will be anonymous, and your child's name, their teachers name, or their school name will not be included in the report.

If you are happy for your child to complete the questionnaires in class, please sign the consent form below and return this to your child's class teacher as soon as possible.

If you change your mind later, you can withdraw your consent without having to give a reason. Your child will also be reminded of their right to withdraw before and after completing the questionnaires.

Please do not hesitate to contact me if you have any questions about the research or would like more information.

Yours Faithfully,

Rachel Faizey

Trainee Educational Psychologist

XXXXX Educational Psychology Service

rachel.faizey@nottingham.ac.uk

Consent form for Parents

School Name:

Class Teacher:

I confirm I have read the information sheet provided and I give permission for_____ **(child's name)** to complete three short questionnaires in the Autumn term to help evaluate Coopertaive Learning approaches used within the school.

Parent/Carer Name:

_____

Parent/Carer Signature

_____

Date:

_____

Please return this consent slip to your child's class teacher by {insert date}

Appendix H Example of Teacher Resource Booklet

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Cooperative Learning



Teacher Resource Booklet

Produce by: Rachel Faizey

(Trainee Educational Psychologist)

Together Everyone Achieves More

(Kagan & Kagan, 2009)

The Principles of Coopertaive Learning - P.I.E.S

Kagan and Kagan (2009) redeveloped the principles of Coopertaive Learning creating 4 key principle using the acronym P.I.E.S.

"When all the principles are in place, we can be sure we will get academic and social gains - if any of the PIES principles are not implemented academic and social gains cannot be guaranteed" (Kagan, 2009).

Positive Interdependence - The group members need each other to complete the task, they cannot succeed unless others do

Individual Accountability - Everyone has a stake in the final outcome and are individual accountable for their own contributions

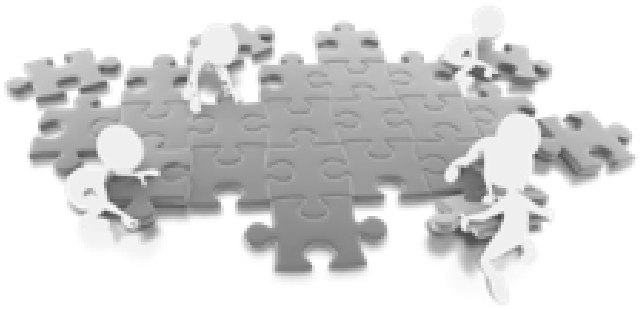
Equal Participation - Everyone in the group makes an equal contribution

Simultaneous Interaction - As many children as possible are actively involved

How do you know all the PIES are covered?

Ask yourself the following questions to make sure you have all the PIES covered:

- Are students on the same side?
- Does the task require working together?
- Is individual, public performance required?
- Is participation approximately equal?
- What percent of students are overly interacting at once?



Social Skills

"People do not know instinctively how to interact effectively with others" (Johnson and Johnson, 1989 p 30)

Social Skills to focus on:

1. Active Listening Skills
2. Asking for Help
3. Disagreeing appropriately
4. Giving and accepting compliments and praise
5. Sharing
6. Taking Turns
7. Talking Quietly
8. Criticising an idea not a person
9. Encouraging others
10. Offering help / coaching

Teaching the Social Skills

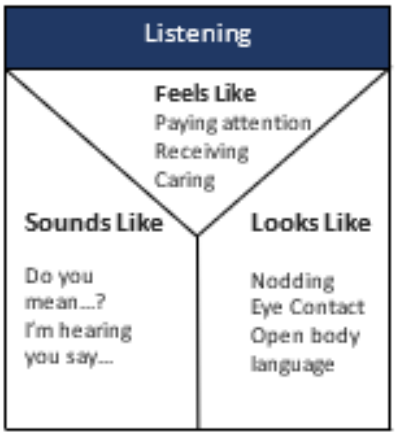
A series of steps can be followed when teaching small group skills.

1. Students seeing the need to use the skill

Students must believe that they will be better of if they know it – highlight the need for the skill by explaining why it is important, displaying what it looks like and informed students they will be rewarded for using it.

2. Students most understand what the skills is and when it should be used, which is most effectively conveyed through modelling

This information is most commonly conveyed through a "Y chart" and through modelling the skill. List the skill and the ask the class "what would this skill look like". After several non-verbal behaviours are generated, the teacher asks "what would this skill sound like?" several example phrases are listed. The teacher then models the skill until the students have a clear idea of what the skills sounds and looks like.



A Y chart to convey what the skill is and what it looks like

3. Students must practice the skill again and again.

After defining the skill, the teacher should ask students to role-play the skill several times with the persons sitting next to them. The social skill may be assigned to students as a role to be engaged in during group tasks. At the end of each cooperative lesson, teachers can announce how many times the skill was observed. New skills need to be cued consistently and reinforced for some time.



Cooperative Learning Structure 1

Timed Pair Share

Timed Think Pair Share

Relevant Social Skills

Active Listening, disagreeing appropriately, encouraging others, Accepting, and giving compliments and praise, Talking Quietly, Sharing and Taking Turns.

In pairs students share with a partner for a predetermined time while the partner listens. Then partners switch roles.

1. Randomly allocate children to pairs (be creative)
2. Teacher announces a topic, states how long each student will share, and provides think time directing children to think about the question.
3. Instruct the children to face their learning partners, work together, share ideas, discuss, clarify and challenge.
4. Partner A shares: Partner B listens
5. Partner B responds with a positive reflection E.G "you were interesting to listen to"
6. Partners switch roles
7. Each partner then shares their ideas with another pair or with the while class. Each child needs to be able to share their partners ideas as well as their own.

Adapted from Kagan and Kagan (2015)

What about the PIES!!!

Positive interdependence

Children are able to learn from each other

Individual accountability

Children need to be accountable to each other for sharing ideas and need to be able to share their partners ideas to another pair of whole group

Equal participation

Each student has an equal opportunity (time) to share their ideas. This can be monitored by the teacher.

Simultaneous Interaction

Each student will be involved in interaction ether through speaking or listening.

Positive Reflections to use:

- Thanks for sharing
- You are interesting to listen to
- "One thing I learned listening to you was…"
- "I enjoyed listening to you because…."
- "Your most interesting idea was…"



Cooperative Learning Structure 2

Numbered Heads Together

Numbered Heads Together

Relevant Social Skills

Active Listening, asking for help, criticizing an idea not a person, encouraging others, offering help, coaching, praising, talking quietly, sharing

Children put their "heads together" to reach an agreement on the team's answer. Everyone needs to keep on their toes because their number may be called to share the team's answers

Set up: Teacher prepares questions or problems to ask teams.

1. Put the children into groups which are of mixed gender and ability (i.e. high, medium high, medium low and low).

2. Number children 1-4

3. Teacher poses a problem or question to the children (e.g. how are rainbows formed?)

4. Pause to give children time to think individually

5. Children then "put their heads together" to talk about their ideas, showing answers, discussing and helping each other.

6. Children sit down when everyone knows the answer or has something to share

7. Teacher calls a number (1-4), and children with that number take it in turns to share their groups answers.

What about the PIES!!!

Positive interdependence

Students can learn from each other, they must work together to ensure they have an agreed answer and check everyone can understand and answer the question

Individual accountability

Students are accountable for to each other for sharing ideas. Everyone must be able to answer the question because no one which number will be called.

Equal participation

By randomly selecting a number all children have an equal chance of taking part

Simultaneous Interaction

All of the children are actively involved in the activity

ALTERNATIVE: Students with numbers could all simultaneously share answers for example using:

- Response cards
- Finger responses
- Chalkboard responses

Appendix I Coaching Session Structure

Coaching Session 1			
EPL component	**How enacted**	**Coaching Session 1: 04.10.2018**	**Feedback from Session**
Relatedness: Feelings of connectedness and social interactions between members of a group.	Opportunities for discussion Establishing teams/learning communities On-site support visits Problem solving activities Group Work	• Initial discussion on: • Implementation of structures • What's working well? • Any questions • Arrange Fidelity Checks • Problem solving – time to think about when and how the structures could be implements (opportunities for group problem solving)	
Authenticity: Genuine and authentic applications of practical content in the setting	Assessment of current practice/current status Consider contextual factors Skills Practice Case Studies One site training and support	• What structures have you tried? What do you want to try next? How you going to use this. • Opportunity to practice and watch videos about the structures. • Opportunity to discuss contextual factors which may interfere with the implementation of the structures	

Autonomy: Consultee-directed learning through opportunities to make meaningful choices and influence the learning experience.	Active Engagement ins setting goals Build autonomy through successful practice Active decision making and problem solving	• Set goals for the next coaching session
Responsiveness: Sustained efforts to assess, attend to and meet consultee needs	Needs and formative assessments guide process Flexible evolving agenda Link consultee knowledge with appropriate experience	• Explore teacher's knowledge about CL, the structures and recap over some of the key principles if needed.
Equity: Valuing all participants as having equal importance and equal power	Making "experts of all participants" Consultee accomplishments	• Make clear the importance of teachers having extensive knowledge

Appendix J Intervention Fidelity Checklist

Date: Intervention/ Control

School: CL Structure used:

Teacher:

Number of Children:

Lesson: Topic

Variable	Check	Present (Yes/No)			Comments
Group	Appropriate group size (no more than 4)	Present (2)	Some Evidence (1)	Absent (0)	
	Heterogenous Groups	Present (2)	Some Evidence (1)	Absent (0)	
	Groups Selected by:	Teacher	Pupil	Random	
Environment	Groups members can see each other	Present (2)	Some Evidence (1)	Absent (0)	
	Each Group has sufficient space	Present (2)	Some Evidence (1)	Absent (0)	
	Necessary materials are readily available	Present (2)	Some Evidence (1)	Absent (0)	
Staff	The task is clearly explained at the outset	Present (2)	Some Evidence (1)	Absent (0)	

	Goals and objectives are explicitly defined	Present (2)	Some Evidence (1)	Absent (0)	
	The necessary small group skills are described	Present (2)	Some Evidence (1)	Absent (0)	
	Staff assist and guide where necessary	Present (2)	Some Evidence (1)	Absent (0)	
Task	Positive Interdependence	Present (4)	Some Evidence (2)	Absent (0)	
	Individual Accountability	Present (4)	Some Evidence (2)	Absent (0)	
	Equal Participation	Present (4)	Some Evidence (2)	Absent (0)	
	Small group skills described	Present (4)	Some Evidence (2)	Absent (0)	
	Reflection on small group skills	Present (4)	Some Evidence (2)	Absent (0)	

Child	Children engaging in Simultaneous Interaction	Present (2)	Some Evidence (1)	Absent (0)	
	Follow instructions accurately	Present (2)	Some Evidence (1)	Absent (0)	
	Can solve conflict independently	Present (2)	Some Evidence (1)	Absent (0)	

Adapted from Craig (2010) and Smith (2017)

Notes….

.

Appendix K Social Inclusion Survey

Questionnaire 1

(Adapted *from the Social Inclusion Survey (Frederickson & Furnham, 1998)).*

How much do you like to **Work** with each person at school?

	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹

How much do you like to **Play** with each person at school?

	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹
	? ☺ 😐 ☹

Appendix L My Life in Schools Checklist

Questionnaire 2 (Adapted from My Life in Schools Checklist, (Arora & Thompson, 1987))

During this week another pupil:

	Not at all	Only Once	More than Once
1. Called me names	☐	☐	☐
2. Tried To kick me	☐	☐	☐
3. Was very nice to me	☐	☐	☐
4. Said they'd beat me up	☐	☐	☐
5. Tried to make me give them money	☐	☐	☐
6. Stopped me playing a game	☐	☐	☐
7. Was unkind about something I did	☐	☐	☐
8. Tried to get me into trouble	☐	☐	☐
9. Helped me carry something	☐	☐	☐
10. Tried to hurt me	☐	☐	☐
11. Helped me with my classwork	☐	☐	☐
12. Took something off me	☐	☐	☐
13. Shared something with me	☐	☐	☐
14. Tried to break something of mine	☐	☐	☐
15. Tried to hit me	☐	☐	☐

Appendix M Empathy Measure

Questionnaire 3 *(Adapted from Emotional Literacy Scales,* (Faupel, 2003)*)*

		Very like me	Quite like me	Only a bit like me	Not like me at all
1	I try to help people when they are unhappy				
2	I carry on trying even if I find something difficult				
3	I find it difficult to make new friends				
4	I am usually included in other children's games				
5	I know when people are starting to get upset				
6	I try hard with my school work				
7	I laugh at other children when they get something wrong				
8	I am good at many things				
9	I try to help someone who is being bullied				

Appendix N Ethics amendment letter

School of Psychology
University of Nottingham
University Park
NG7 2RD
Supervisor: Anthea Gulliford
14th September 2018

To the Chair of the Ethics Committee,

Ethics Approval Reference: S1078

Following the ethical approval received for the proposed research **"An Evaluation of the impact of a Cooperative Learning intervention on PA, empathy and pupils experience of bullying in school"** (reference S1078) in the letter dated 17th May 2018. I would like to request ethical approval for an amendment in the research methodology.

I would like to make an amendment to the Empathy measure used. The measure outline in the proposal was the Kids Empathy Developmental Scale (KEDS) (Reid et al, 2013). Piloting revealed problems with the measure: it took a significant amount of time to administer, would be difficult to complete with a whole class and used images which did not seem to reflect current social contexts.

Therefore, I would like to use an alternative measure and after further exploring and comparing empathy measures available I propose using the empathy factor from the Emotional Literacy Scales pupil checklist (Faupel, 2003). The Emotional Literacy Scales pupil checklist is a widely used self-report measure, which looks aspects of personal and social competence including empathy and social skills, for students aged 7-11. It was standardised and has reported reliability overall of 0.76. The reliability for the empathy dimension is lower. The following steps will therefore be taken:

- Reliability coefficients will be recalculated within this study
- Descriptive data will be capture only if need, for the empathy dimension

- Triangulation between this and other measures
- A full debate upon the lack of well-attuned measures of empathy in middle childhood will be presented in the thesis report

The adapted scale I would like to use can be found in Appendix 1 and contains the 5 items from the Emotional Literacy Scales related to Empathy and 5 distraction items which will not be scored.

Yours Faithfully,

Rachel Faizey

Appendix O Pupil Information

Miss Faizey

Trainee Educational Psychologist

Why am I here?

Your school and teachers have been trying some new ways of organising your learning.

Some of the teachers in your school have been trying these new ways from September.

We need your help to see if these new ways of organising your learning are useful or not.



What do I need to do?

- You will be asked to fill in 3 questionnaires again just like the ones you did in September.
  - **Questionnaire 1** will ask you about who you like to work and play with in your class
  - **Questionnaire 2** will ask you about your experience of friendly or unfriendly behaviour in school.
  - **Questionnaire 3** will ask what you think about yourself in school

What is a questionnaire?

A questionnaire is a list of questions which help us know what people think



Questionnaire 1

We would like to know how much you like to **Work** which each person at school?

? Someone you don't know well enough to decide

☺ Name of someone you like to work with

😐 Someone you don't mind working with

☹ Someone you prefer not to work with

Questionnaire 3 (Adapted *from the Social Inclusion Survey (SIS (Fredrickson & Furnham, 1998)*

How much do you like to **Work** with each person at school?

	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹

Questionnaire 1

We would like to know how much you like to **Play** which each person at school?

? Someone you don't know well enough to decide

☺ Name of someone you like to work with

😐 Someone you don't mind working with

☹ Someone you prefer not to work with

How much do you like to **Play** with each person at school?

	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹
	?	☺	😐	☹

Questionnaire 2

We would like to know what happens to people in school.

On this page there are various statements about things that might have happened to you during the last week.

- If they have not happened during the last week tick "Not at all"
- If they have happened once tick "Only once"
- If they have happened more than once this week tick "More than once"

Questionnaire 1 (Adapted from My Life in Schools Checklist)

During this week another pupil:

	Not at all	Only Once	More than Once
1. Called me names	☐	☑	☐
2. Tried To kick me	☑	☐	☐
3. Was very nice to me	☐	☐	☑
4. Said they'd beat me up	☐	☐	☐
5. Tried to make me give them money	☐	☐	☐
6. Stopped me playing a game	☐	☐	☐
7. Was unkind about something I did	☐	☐	☐
8. Tried to get me into trouble	☐	☐	☐
9. Helped me carry something	☐	☐	☐

Questionnaire 3

We would like to know what you think about yourself in school

Questionnaire 3 ***(Adapted from Emotional Literacy Scales)***

		Very like me	Quite like me	Only a bit like me	Not like me at all
1	I try to help people when they are unhappy		✓		
2	I carry on trying even if I find something difficult			✓	
3	I find it difficult to make new friends				
4	I am usually included in other children's games				
5	I know when people are starting to get upset				
6	I try hard with my school work				
7	I laugh at other children when they get something wrong				
8	I am good at many things				
9	I try to help someone who is being bullied				

Who will see my answers?

- Your questionnaires will be collected in and your answers will be kept private and safe.
- Your name will not be on the questionnaire and this means your teachers, parents or any other children will not know what answers you have put.
- **This means you can answer the questions honestly**



What will happen next?

- I will look at everyone's questionnaires to see if whether the different ways or organising your learning are useful or not.
- I will write a report to tell people what I have found.
- Don't worry this will not name you, your friends, your teacher or your school.

What do I do now?

- Put your age and tick whether you are a boy or a girl
- On the next page tick whether you are happy to do the questionnaires

Pre - Questionnaires

Age:

Date:

Boy ☐ Girl ☐

Do I have to fill in these questionnaires?

- Your Parents/Carers have said they are happy for you to fill in the questionnaires.
- But if you don't want to fill out the questionnaires you can say no.
- You will see a page that looks like this and on here you can tick whether you happy to fill in the questionnaires.

Cooperative Learning
Pupil Consent Form

- I have read/had the information sheet for children read to me.
- I understand I am going to be given three questionnaires to help the researcher find out if new ways of organising learning can be useful.
- I understand I will be asked how much I like to work and play with other children in my class, asked to look at pictures and think about how people in the pictures might be feeling and asked about my experience of school
- I understand that I can choose whether to fill in the questionnaires or not, and no one will be cross if I say no.
- I understand that I can change my mind later and ask for my answers to not be included.
- I understand that all my answers will be kept private and will not be shown to anyone
- I have been able to ask any questions

Please tick one of the following boxes: ☑

I am happy to fill in the questionnaires ☐

I do not want to fill in the questionnaires ☐

Any Questions?



Before we begin….

- Listen to the instructions
- There are no right or wrong answers
- Keep your answers private - remember we want you know what you think

Appendix P Pupil Consent Form

Name:

Date:

School:

Teacher:

- I have read/had the information sheet for children read to me.
- I understand I am going to be given three questionnaires to help the researcher find out if new ways of organising learning can be useful.
- I understand I will be asked how much I like to work and play with other children in my class, asked to look at pictures and think about how people in the pictures might be feeling and asked about my experience of school
- I understand that I can choose whether to fill in the questionnaires or not, and no one will be cross if I say no.
- I understand that I can change my mind later and ask for my answers to not be included.
- I understand that all my answers will be kept private and will not be shown to anyone
- I have been able to ask any questions

Please tick one of the following boxes: 

I am happy to fill in the questionnaires 

I do not want to fill in the questionnaires 

Appendix Q Ethical approval from the University of Nottingham Ethics Committee

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

School of Psychology
The University of Nottingham
University Park
Nottingham
NG7 2RD

tel: +44 (0)115 846 7403 or (0)115 951 4344

SJ/tp

Ref: S1078

Thursday 17th May 2018

Dear, Anthea and Rachel,

Ethics Committee Review

Thank you for submitting an account of your proposed research **"An evaluation of the impact of a Cooperative Learning intervention on peer acceptance, empathy and pupils experience of bullying in school"**

That proposal has now been reviewed and we are pleased to tell you it has met with the Committee's approval.

However:
Please note the following comments from our reviewers;

Reviewer One:

- I would recommend including in the parent information sheet that children in the control group will also be given the CL activities upon completion of the study, especially if this intervention is found to be beneficial for pupils.

Reviewer Two:

- Typo in the letter "Access and recruitment letter – parents PILOT". "...your child to be in the **date** collection element of the study", "date" should be "**data".**
- Wait list control group. In the letters to schools and parents is not mentioned that the wait list control group will be offered the same training as the experiment group following the intervention. I believe it is important to mention this in the letters.

Final responsibility for ethical conduct of your research rests with you or your supervisor. The Codes of Practice setting out these responsibilities have been published by the British Psychological Society and the University Research Ethics Committee. If you have any concerns whatever during the conduct of your research then you should consult those Codes of Practice. The Committee should be informed immediately should any participant complaints or adverse events arise during the study.

Independently of the Ethics Committee procedures, supervisors also have responsibilities for the risk assessment of projects as detailed in the safety pages of the University web site. Ethics Committee approval does not alter, replace, or remove those responsibilities, nor does it certify that they have been met.

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

School of Psychology
The University of Nottingham
University Park
Nottingham
NG7 2RD

tel: +44 (0)115 846 7403 or (0)115 951 4344

Yours sincerely



Professor Stephen Jackson
Chair, Ethics Committee

Appendix R Parent debriefing letter



Dear Parent/Carer

Thank you for giving your permission for your child to take part in my research into Cooperative Learning which is an intervention approach in which students work together to achieve shared learning goals and aims to promote learning and social skills

The aim of the research was to evaluate what effect Cooperative Learning has on levels of PA, empathy, and pupil's experiences of friendly and bullying behaviour in schools.

Your child was asked to complete three questionnaires in the classroom (on two occasions) to evaluate how useful Cooperative Learning has been. The information your child gave me will be held anonymously. This means that no one will know what answers they have given.

One of the questionnaires asked about friendly and unfriendly behaviour your child may have experienced in school. If your child has experienced unfriendly behaviour in school and you feel they would like to talk about this, please speak to or ask them to speak to (insert name of school staff member).

If you any questions you would like to ask me, please do not hesitate to contact me on the details provided above.

Yours Sincerely,

Rachel Faizey

Trainee Educational Psychologist

rachel.faizey@nottingham.ac.uk

Debrief Information for pupils

As you know some of your teachers in your school were trying new ways of organising your learning. The aim was to see if these new ways of learning helped people in your class work together better. You were asked to fill in three questionnaires which helped us to see how these new ways of learning helped pupils work together.

I have taken your names off the questionnaires, so nobody will know they are your questionnaires and they will be kept safe.

One of the questionnaires asked about friendly and unfriendly behaviour you may have experienced in school. If you feel you have experienced unfriendly behaviour in school and would like to talk about this, please speak to (insert name of school staff member).

If you do not wish your answers to be included, please let (insert name of staff) know by **15th January 2019.**

If you think of any questions you would like to ask once I have gone then you can speak to (insert name of school staff member).

Thank you for help.

Rachel Faizey

Trainee Educational Psychologist

Appendix S Examples of completed teacher record booklets

Date	Day	Structure Employed Number of children working together	Time spent using CL structure	Curriculum Area	Children who were absent for the session	Additional Comments
	Mon					
	Tues	T-P-S	2 mins	Topic	[redacted]	Nice way to end
		30→15				lesson. Boring comp
	Weds	TP-S	5 mins	Languages		more independently.
		30→15				
	Thurs	Rally Robin	2 mins	x tables	—	Quick and effective
	Fri	//	/	/	/	/

Date	Day	Structure Employed Number of children working together	Time spent using CL structure	Curriculum Area	Children who were absent for the session	Additional Comments
24/09	Mon	Think pair share	8 mins	PSHE	[redacted]	were excited to start task and liked active listening
		20 - children in pairs				
25/09	Tues	Rally Robin	10 mins	Topic	[redacted]	Interesting to see who was listening and their active listening skills better that last time
		20 - In pairs				
26/09	Weds	Think Pair share	5 mins	Topic	[redacted]	
27/09	Thurs	Quiz Quiz trade	12 mins	"	—	
	Fri					

Appendix T Bullying Index Data

Class	**Intervention**	**Percentage of pupils that ticked 'more than once'**	**Item 1**	**Item 2**	**Item 3**	**Item 4**	**Item 5**	**Item 6**	**Bullying Index**
School A Year 5	CL	Pre-Test	5.26	21.05	5.26	21.05	15.79	21.05	**14.91**
		Post-test	9.09	9.09	0.00	13.64	5.882	5.56	**7.21**
	WLC	Pre-Test	26.32	21.05	10.53	26.32	10.53	31.58	**21.05**
		Post-test	21.05	10.53	10.53	26.32	5.556	21.05	**15.84**
School B Year 4	CL	Pre-Test	9.09	18.18	4.55	13.64	4.55	13.64	**10.61**
		Post-test	0.00	0.00	5.26	5.00	0.00	5.00	**2.54**
	WLC	Pre-Test	15.79	0.00	0.00	15.79	5.56	16.67	**8.97**
		Post-test	16.67	5.56	11.11	22.22	11.11	11.11	**12.96**

School B Year 5	CL	Pre-Test	4.35	4.35	0.00	13.64	9.09	4.55	**5.99**
		Post-test	31.58	5.26	5.26	36.84	16.67	31.58	**21.20**
	WLC	Pre-Test	19.05	5.00	0.00	14.29	5.00	15.79	**9.85**
		Post-test	35.00	21.05	10.00	30.00	15.00	40.00	**25.18**

Item 1 = Tried to kick me

Item 2 = Said they'd beat me up

Item 3 = Tried to make me give them money

Item 4 = Tried to hurt me

Item 5 = Tried to break something of mine

Item 6 =Tried to hit me

Appendix U General Aggression Index Data

Class	**Intervention**	**Percentage of pupils that ticked 'only once'**	**Item 1**	**Item 2**	**Item 3**	**Item 4**	**Item 5**	**Item 6**	**General Aggression Index**
School A Year 5	CL	Pre-Test	26.32	5.26	10.53	15.79	15.79	5.26	**14.04**
		Post-test	13.64	4.55	9.09	4.55	0.00	11.11	**7.18**
	WLC	Pre-Test	42.1	15.79	5.26	26.32	21.05	5.26	**20.18**
		Post-test	10.53	10.53	0.00	10.53	0.00	10.52	**11.43**
School B Year 4	CL	Pre-Test	13.64	0.00	9.09	13.64	0.00	13.64	**9.47**
		Post-test	10.53	10.53	5.26	5.00	0.00	5.00	**4.30**
	WLC	Pre-Test	15.79	5.26	5.26	15.79	11.11	27.78	**11.23**
		Post-test	0.00	0.00	5.56	22.22	5.56	11.11	**4.30**

School B Year 5	CL	Pre-Test	26.09	13.04	0.00	40.91	0.00	31.82	**12.32**
		Post-test	15.79	35.84	5.26	15.79	11.11	31.58	**20.30**
	WLC	Pre-Test	19.05	10.00	0.00	14.29	0.00	5.26	**8.98**
		Post-test	10.00	10.53	5.00	50.00	5.00	25.00	**21.38**

Item 1 = Tried to kick me

Item 2 = Said they'd beat me up

Item 3 = Tried to make me give them money

Item 4 = Tried to hurt me

Item 5 = Tried to break something of mine

Item 6 =Tried to hit me